POL 138 Quantitative Reasoning in Political Science
KEY Practice Exam 2

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 2 should closely match content on Practice Exam 2, 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.

5 The problem of alternate explanations

5.1 Correlations

1. Of the following, which term is most appropriate to describe a measure of the extent to which the values of one variable associate with the values of another variable?
   A. a correlation
   B. an inference
   C. a percentile
   D. a standard deviation

2. If the numbers in X increase as the numbers in Y increase, then that is a ___ between X and Y.
   A. positive correlation
   B. negative correlation
   C. zero correlation

   For positively correlated variables, higher values of one variable tend to associate with higher values of the other variable (or you can think of it as lower values of one variable tending to associate with lower values of the other variable).

3. If the numbers in X increase as the numbers in Y decrease, then that is a ___ between X and Y.
   A. positive correlation
   B. negative correlation
   C. zero correlation

   For negatively correlated variables, higher values of one variable tend to associate with lower values of the other variable.

4. If the numbers in X decrease as the numbers in Y decrease, then that is a ___ between X and Y.
   A. positive correlation
   B. negative correlation
   C. zero correlation

For positively correlated variables, higher values of one variable tend to associate with higher values of the other variable (or you can think of it as lower values of one variable tending to associate with lower values of the other variable).

5. If the numbers in X do not change as the numbers in Y increase, then that is a ___ between X and Y.
   A. positive correlation
   B. negative correlation
   C. zero correlation

   A zero correlation is when the variables do not associate with each other.

5.2 Alternate explanations

6. Suppose that data indicated that political knowledge was higher on average among political science majors than among education majors, with a p-value of $p<0.05$ for a test of the null hypothesis that these means equal each other. Is this sufficient evidence to conclude at the conventional level in political science that being a political science major caused a higher level of political knowledge than being an education major did, at least on average?
   A. Yes, because the p-value is $p<0.05$, and it makes sense that political science classes would cause higher levels of political knowledge than education classes would.
   B. No, because the analysis did not address alternate explanations such as the possibility that, even before these students entered their majors, political knowledge was higher among students who planned to major in political science than among students who planned to major in education.

7. Suppose that data from a large nationally representative sample of U.S. residents indicated that U.S. residents who reported being sexually harassed at work were more likely to report being a Democrat than to report being a Republican. Explain whether this is sufficient evidence to conclude that, at least on average and among this sample of U.S. residents, being sexually harassed at work caused a person to be more likely to report being a Democrat than to report being a Republican.

   No, because there are plausible alternate explanations that should first be addressed. For example, women are more likely to be Democrats than to be Republicans, and -- if women are more likely to be sexually harassed than men are -- that might explain why U.S. residents who reported being sexually harassed at work were more likely to report being a Democrat than to report being a Republican.

8. The SAT is a test that some colleges use to determine whether to admit a student. Some states require all students in that state to take the SAT during school on an "SAT School Day", and the state pays for all students to take the SAT. But some states don't require any students to take the SAT, although, in these states, students are permitted to take the SAT, and these students often take the SAT on the weekend and pay for the SAT on their

own. Data for the SAT in 2022 indicated that the mean SAT math score was 577 for students who took the SAT on the weekend, but was 451 for students who took the SAT on SAT School Day. The p-value is $p<0.01$ for a test of the null hypothesis that these scores equal each other. Discuss whether this is sufficient evidence at the conventional level in political science that, compared to taking the SAT on SAT School Day, taking the SAT on the weekend caused students to do better on the SAT, at least on average.

A plausible expectation is that requiring all students to take the SAT reduces the mean SAT score, because a lot of the students who would not have taken the SAT if the SAT were optional are not planning to go to college, and part of the reason for not going to college is that some of these students have not done well enough academically in high school to make college a good decision. So the students who take the SAT on SAT School Day are plausibly on average not as academically good as the students who take the SAT on the weekend. So, given the plausible chance that the SAT test-takers on SAT School Day differ on average academically from SAT test-takers on the weekend, the data reported in the item is not sufficient evidence that taking the SAT on the weekend caused students to do better on the SAT, at least on average.

9. Data from our POL 138 class indicated that, on average, the number of class meetings a student attended positively associated with the student's score on Exam 2, with a p-value of $p<0.05$ for a test of the null hypothesis that the number of class meetings a student attended did not associate with the student's score on Exam 2. It is possible that this positive association was because students attending more class meetings on average caused students to score higher on Exam 2, because students learned while in class. But provide a different plausible reason why the number of class meetings a student attended positively associated with the student's score on Exam 2.

There are many acceptable responses for this item. For example, maybe the type of student who attended class meetings more often was also the type of student who was more likely to read the course notes on their own or to study more or to go to tutoring...and maybe these other things caused that type of student to do better on Exam 2.

6 Randomized experiments

6.1 Randomized experiments

10. Which one of these is NOT a necessary step in a randomized experiment involving human participants?
    A. Treat each group differently.
    B. Randomly assign participants to groups.
    C. Use control variables to eliminate alternate explanations.
    D. Measure some outcome for each group.

11. Randomly assigning participants to groups helps a randomized experiment identify causes by ___.
    A. eliminating demand effects as much as possible
    B. getting the groups to be as similar to each other as possible before the difference in treatment
    C. getting the sample to be as representative of the population as possible without weighting
    D. helping as much as possible to avoid regression toward the mean

12. 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. 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

13. Random assignment error in a randomized experiment ___.
    A. can bias an estimate of an effect only to be lower than it truly is
    B. can bias an estimate of an effect only to be higher than it truly is
    C. can bias an estimate of an effect to be lower than or higher than it truly is
    D. cannot bias an estimate

14. Suppose that a researcher conducted a randomized experiment and then compared the mean response from participants in the control group to the mean response from participants in the treatment group. The p-value was p=0.01 for a test of the null hypothesis that these means equal each other. Based on this p-value, the researcher should conclude that ___.
    A. the treatment had an effect
    B. the treatment did not have an effect
    C. there is not enough evidence to conclude that the treatment had an effect

15. Suppose that a researcher conducted a randomized experiment and then compared the mean response from participants in the control group to the mean response from participants in the treatment group. The p-value was p=0.25 for a test of the null hypothesis that these means equal each other. Based on this p-value, the researcher should conclude that ___.
    A. the treatment had an effect
    B. the treatment did not have an effect
    C. there is not enough evidence to conclude that the treatment had an effect

[Items 16 and 17] A researcher randomly selects 200 people from a population and then randomly assigns 100 of these people to a group that receives Treatment A and randomly assigns the other 100 people to a group that receives Treatment B.

16. The random assignment to groups ___.
    A. better permits the researcher to make an inference about the population
    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

17. The random selection from the population ___.
    A. better permits the researcher to make an inference about the population
    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

---

18. Suppose that researchers have a sample in which 50 persons are randomly assigned to watch Video A and 50 persons are randomly assigned to watch Video B. Both videos encourage people to donate blood, and the only difference between the videos is that Video A ends with the narrator saying "Please donate blood" and Video B ends with the narrator saying "Please donate blood, for the children". After watching the video, each participant is asked to donate blood.

    Suppose that exactly 10 of the 50 persons who watched Video A donated blood after being asked to donate blood (20%) and that exactly 20 of the 50 persons who watched Video B donated blood after being asked to donate blood (40%). The p-value is $p<0.05$ for a test of the null hypothesis that these percentages equal each other. Explain whether this is sufficient evidence at the conventional level in political science to conclude that the "for the children" at the end of Video B caused the difference between groups in the percentage of persons who donated blood.

    Yes, in a randomized experiment, the only two reasons for a difference between the groups is [1] random assignment error or [2] the difference in treatment. The p-value under p=0.05 permits us to rule random assignment error as a plausible reason for the difference between groups, so the only remaining reason is the difference in treatment.

    For this item, some students in the past have responded "No", noting that the difference between groups could have been caused by random assignment error. That is true, but the item statement asked about whether this was "sufficient evidence at the conventional level in political science", so the p=0.01 p-value was enough to eliminate random assignment error as a plausible explanation for the difference between groups, at the conventional level in political science.

    It's true that random assignment error could have caused (for instance) a higher percentage of compassionate people to be assigned to Video B than to Video A, but the p-value under p=0.05 lets us rule out that type of random assignment error as a

plausible explanation for the difference between groups, at the conventional level in political science.

19. Suppose that, in a correctly conducted 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. Indicate 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.

The other possible reason is that random assignment error caused the difference. Random assignment error refers to differences between the groups after the randomization but before the difference in treatment. So, for instance, if random assignment produced a control group that was 51% female and a treatment group that was 49% female, that 2% difference would be due to random assignment error.

20. Randomly assigning participants to groups in an experiment helps to reduce a certain kind of bias that might occur if participants were able to select whether they wanted to be in the control group or the treatment group. Explain how randomly assigning participants to groups helps eliminate this bias.

Randomization helps ensure (as much as we can) that the groups will be similar to each other on all characteristics before the difference in treatment. If participants grouped on their own, the participants might group based on similar characteristics such as race or gender, which would make the groups very unequal before the difference in treatment.

21. Bob wants to test whether a pill causes weight loss, so he assigns a randomly selected set of 1,000 U.S. residents to take the pill each morning for ten weeks. Results indicated that the mean weight of the participants decreased over the ten weeks of the study ($p<0.001$). Based on this evidence, can we conclude at the conventional level in political science that the pill caused the weight loss among these participants, at least on average?
A. Yes
B. No

6.2 Placebos

22. 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

6.3 Natural experiments

23. Which one of the following indicates a difference between a randomized experiment and a natural experiment?
    A. In a natural experiment, the experiment must be conducted outside.
    B. In a natural experiment, computers must not be used for the data analysis.
    C. In a natural experiment, the outcome variable must be a measure of a natural phenomenon.
    D. In a natural experiment, the assignment of the treatment must be done by nature or as if by nature.

7 Non-random comparisons

7.1 Discontinuity designs

24. Faber College offers POL 100 each Monday and Wednesday in four sections, with start times of 8am, 11am, 2pm, and 6pm. Each class meeting is 1 hour and 15 minutes long. Lunch at Faber College is from 12:30pm to 1:30pm. Students are randomly assigned to sections of POL 100. For each section, students take a pretest and a posttest to measure their learning over the semester. Researchers are interested in the effect of eating lunch on learning. Researcher A plans to compare the average student learning across the two sections before lunch (8am and 11am) to the average student learning across the two sections after lunch (2pm and 6pm). Researcher B plans to instead compare the average student learning in the 11am section to the average student learning in the 2pm section. An advantage of Researcher B's research design over Researcher A's research design is that ___.
    A. Researcher B will avoid Simpson's paradox
    B. Researcher B will have a smaller sample size
    C. Researcher B will address an alternate explanation
    D. Researcher B will avoid bias due to regression toward the mean

25. In the United States in the 1930s, the U.S. government sponsored the Home Owners' Loan Corporation, which created maps. On these maps, certain areas were colored green, blue, yellow, or red, to indicate the perceived mortgage lending risk in these areas. The areas that were colored red were considered to be the riskiest areas to lend to, and the areas that were colored green were considered the least risky areas to lend to. The process of assigning geographic regions to the red area is called "redlining".

    For areas that were redlined on these maps in the 1930s, economic outcomes are on average relatively poor in modern times: for example, in 2016, the percentage of residents who were low-to-moderate income was 9% for the green areas, but 74% for the red areas. The p-value is $p<0.05$ for a test of the null hypothesis that the modern-day percentage of residents in the green areas who are low-to-moderate income equals the modern-day percentage of residents in the red areas who are low-to-moderate income. Explain whether this is sufficient evidence at the conventional level in political science

that the "redlining" maps in the 1930s caused this modern-day difference between residents in the green areas and residents in the red areas in the percentage of persons who are low-to-moderate income.

No, because there are plausible alternate explanations. For example, areas that were redlined might have already had poor economic outcomes on average in the 1930s, so that the causal direction is instead that these poor economic outcomes caused the redlining, instead of the other way around.

26. Regarding the redlining discussed in the prior item, Aaronson et al 2018 reported on an analysis that limited comparisons to edges between colored regions, to, for example, compare [1] outcomes for residents who live in a redlined area but who live near the edge of that redlined area to [2] outcomes for residents who live right across the street from that redlined resident but who live in a greenlined area.

For the purpose of estimating whether redlining has had a negative effect on modern-day outcomes for people living in redlined areas, explain an advantage of limiting comparisons to the edges between redlined areas and greenlined areas, instead of comparing outcomes for all residents of redlined areas to outcomes for all residents of greenlined areas.

The limited comparison can help address alternate explanations by making the comparison persons more similar to each other on average, except for the color of the map area in which they live. For example, the average redlined person might differ in a lot of way from the average greenlined person in income and wealth and employment, but these differences are plausibly a lot smaller for "redlined" persons who live right across the street from "greenlined" persons.

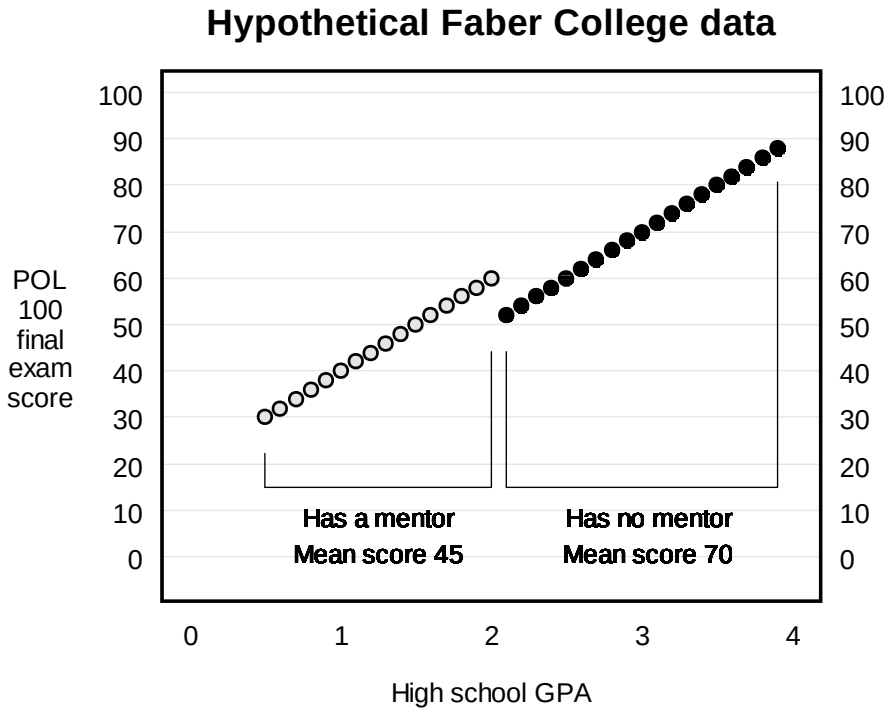
27. Suppose that — based on a student's SAT or ACT score, the student's current GPA, the student's frequency of class attendance, and the rigor of the student's major — a university calculates an "academic risk" score that predicts each student's risk of dropping out of the university. Based on this academic risk score, the university divides its 20,000 students into 100 groups of 200 students each, in which students in Group 1 have the highest academic risk score, which means that these students are predicted to have the highest risk of dropping out; students in Group 2 have the next highest academic risk score, and so on, with students in Group 100 having the lowest academic risk index score, which means that these students are predicted to have the lowest risk of dropping out. The university provides to each student in Groups 1 through 50 — and only to these students — a University Academic Mentor who communicates with the student at least once per week and provides other academic support.

The university is interested in determining a research design for assessing the effect of the University Academic Mentor on student dropout rates. Researcher A suggests that the dropout rate for students in Groups 1 through 50 (in which all students were assigned a University Academic Mentor) be compared to the dropout rate for students in Groups 51 through 100 (in which no student was assigned a University Academic Mentor).

Researcher B suggests that the dropout rate for students in Groups 49 and 50 be compared to the dropout rate for students in Groups 51 and 52. Explain an advantage of Researcher B's research design over Researcher A's research design.

Compared to the two groups that Researcher A proposes to compare, students in Groups 49 and 50 should be closer on average on all relevant traits to students in Groups 51 and 52, so that the major difference between Researcher B's two groups is only the use of the University Academic Mentor.

[Items 28 and 29] Suppose that Faber College has a program that assigns freshman students a mentor if and only if the student had a high school GPA that was 2.0 or lower. Each Faber College freshman must take a POL 100 course. The plot below indicates, for each Faber College freshman, the student's high school GPA, the student's final exam score in POL 100, and whether Faber College assigned the student a mentor.



28. Suppose that we (incorrectly) estimate the effect of having a mentor on a student's POL 100 final exam score, by comparing how much the mean final exam score among students who had a mentor differs from the mean final exam score among students who did not have a mentor. Which of the following best indicates that (incorrect) estimate?

A. The mentor reduced final exam scores by about 25 points, on average.
B. The mentor reduced final exam scores by about 10 points, on average.
C. The mentor did not affect final exam scores, on average.
D. The mentor increased final exam scores by about 10 points, on average.
E. The mentor increased final exam scores by about 25 points, on average.

The "no mentor" mean score of 70 is 25 units higher than the "mentor" mean score of 45.

29. Suppose that we instead use a discontinuity method to estimate the effect of having a mentor on a student's POL 100 final exam score, by comparing how much the final exam score among students who were just below the threshold for getting a mentor differs from the final exam score among students who were just above the threshold for getting a mentor. Which of the following best indicates that (more plausible) estimate?
A. The mentor reduced final exam scores by about 25 points, on average.
B. The mentor reduced final exam scores by about 10 points, on average.
C. The mentor did not affect final exam scores, on average.
D. The mentor increased final exam scores by about 10 points, on average.
E. The mentor increased final exam scores by about 25 points, on average.

Right below the GPA threshold, the "no mentor" score is about 60. Right above the GPA threshold, the "mentor" score is about 50.

7.2 Difference-in-differences designs

30. Suppose that, at Faber College, enrollment in the political science major increased 2% each year from 2012 to 2017. In 2018, the political science department got a new department chair, and, over the next five years, enrollment increased at only 1%. For estimating how the new chair affected enrollment rates in the political science major, which of the following would provide the better comparison for a difference-in-differences design, based on only the enrollment rates indicated below?
A. the history major at Faber College, in which enrollment increased at 1% per year from 2012 to 2017
B. the sociology major at Faber College, in which enrollment increased at 2% per year from 2012 to 2017
C. the economics major at Faber College, in which enrollment increased at 2% per year from 2018 through 2022

A difference-in-differences design includes a comparison group that, before the treatment, was similar to the group of interest and, as best we can tell, should be expected to have been similar to the treated group afterwards, if not for the difference in treatment. So, in this case, we try to match the 2% political science increase from 2012 to 2017 to another similar 2% increase from 2012 to 2017.

[Items 31 and 32] Suppose that a researcher is interested in the extent to which college causes persons to become more politically liberal. In 2019, the researcher surveys a representative sample of age-18 persons who attend college and a representative sample of age-18 persons who do not attend college. Four years later, in 2023, the researcher surveys each person again. Suppose that the researcher's data are in the table below, in which political ideology is measured from 0 for extremely liberal to 10 for extremely conservative.

Group	Mean ideology	Mean ideology

	at age 18	at age 22
Persons in college	4.5	3.5
Persons not in college	5.0	4.2

31. If the researcher analyzed only the data for persons in college, the researcher's (incorrect) estimate of the effect of college on political ideology would be that college ___.
A. made persons in the sample about 0.2 units more liberal on average
B. made persons in the sample about 0.8 units more liberal on average
C. made persons in the sample about 1.0 unit more liberal on average
D. made persons in the sample about 3.5 units more liberal on average

The change among persons in college is from 4.5 to 3.5, which is 1.0 units more liberal.

32. If the researcher used a difference-in-differences design that compared persons in college to persons not in college, the researcher's (more correct) estimate of the effect of college on political ideology would be that college ___.
A. made persons in the sample about 0.2 units more liberal on average
B. made persons in the sample about 0.8 units more liberal on average
C. made persons in the sample about 1.0 unit more liberal on average
D. made persons in the sample about 3.5 units more liberal on average

The change among persons in college is from 4.5 to 3.5, which is 1.0 units more liberal. The change among persons not in college is from 5.0 to 4.2, which is 0.8 units more liberal. So, compared to persons not in college, persons in college got 0.2 units more liberal (calculated as 1.0 more liberal minus 0.8 more liberal).

---

33. 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. Oceania is a country immediately next to Freedonia and is similar to Freedonia in every way, except that Oceania did not enact any legislation to reduce unemployment.

	Unemployment Rate			
	2021	2022	2023	2024
Freedonia	6%	6%	6%	3%
Oceania	6%	6%	6%	3%

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 Unemployment Reduction Act was plausibly the reason for the decrease in unemployment in Freedonia between 2023 and 2024.
B. The Unemployment Reduction Act was probably not the reason for the decrease in unemployment in Freedonia between 2023 and 2024.

The decrease in Freedonia between 2023 and 2024 is the same as the decrease in Oceania between 2023 and 2024, so there does not seem to be anything special about the legislation that was passed in Freedonia.

34. Suppose that unemployment in Freedonia was 12% in 2021, 12% in 2022, and 12% in 2023, and then, on January 1, 2024, Freedonia enacted the Unemployment Reduction Act. Unemployment was then 6% in 2024. Researchers are interested in assessing the extent to which the Unemployment Reduction Act caused this decrease in the unemployment rate. Oceania and Pacifica are countries next to Freedonia and are similar to Freedonia in many ways, but neither Oceania nor Pacifica enacted any legislation to reduce unemployment.

	Unemployment Rate			
	2021	2022	2023	2024
Freedonia	12%	12%	12%	6%
Oceania	14%	14%	14%	14%
Pacifica	10%	10%	10%	10%

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 Unemployment Reduction Act was likely the reason for the decrease in unemployment in Freedonia between 2023 and 2024.

B. The Unemployment Reduction Act was likely not the reason for the decrease in unemployment in Freedonia between 2023 and 2024.

The decrease in Freedonia between 2023 and 2024 is not the same as the decrease in Oceania between 2023 and 2024, so there might be something special about the legislation that was passed in Freedonia.

7.3 Benchmarks

35. 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 students who were in Pi Sigma Alpha to future income among any college students who were not in Pi Sigma Alpha. Researcher B instead plans to compare future income among any college students who were in Pi Sigma Alpha and had a 3.0 GPA to future income among any college students 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 on the future income of college graduates?

A. Researcher A, because Researcher A will have a much larger sample size and thus a more precise estimate of the effect, if any, that membership in Pi Sigma Alpha has 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.

36. Researchers are interested in assessing whether the percentage of women among U.S. political science professors is unfairly too high or too low due to gender bias among the people who hire political science professors. Researcher A plans to compare the percentage of women among U.S. political science professors to the percentage of women among all U.S. residents. Researcher B instead plans to compare the percentage of women among U.S. political science professors to the percentage of women among U.S. residents who have a Ph.D. in political science. Explain why Researcher B's suggested benchmark might be better than Researcher A's suggested benchmark.

People who hire political science professors cannot hire just anyone from the general population but instead must choose from the pool of people who have an advanced degree in political science. The benchmark of the percentage of women among U.S. residents who have a Ph.D. in political science thus better reflects this pool of potential hires.

7.4 Panel designs

37. Suppose that, for two participants, a researcher has data from two surveys, one in April and another 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. Data are below, with each participant identified with an ID:

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

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

In this case, the political party remains the same but the attitudes change, so the presumption is that the political party is influencing the attitudes.

38. Suppose that, for two participants, a researcher has data from a survey in January and another survey in December of the same year, with each participant appearing twice in the dataset. 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. Data are below, with each participant identified with an ID:

ID	January	December

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

In this case, the political party remains changes but the attitudes remain the same, so the presumption is that the attitudes are influencing political party.

8 Statistical control

8.1 Statistical control

39. How do control variables help improve causal inference in a correlational study?
A. reduce bias in measurements
B. remove sample bias
C. help address alternate explanations
D. randomly assign participants

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

41. In a non-experimental analysis, including a control variable that should not be included ___.
A. can bias an estimated effect to be too low but cannot bias an estimate to be too high
B. can bias an estimated effect to be too high but cannot bias an estimate to be too low
C. can bias an estimated effect to be too low and can bias an estimate to be too high
D. cannot bias an estimated effect to be too low and cannot bias an estimate to be too high

42. Suppose that you were asked to conduct a study to determine whether male ISU employees are paid more than female ISU employees are paid. For each ISU employee, you have data on the employee's salary and their gender. Using these salary data as the outcome, would you need control variables for this study?
A. Yes
B. No

Determining whether male ISU employees are paid more than female ISU employees are paid is a descriptive question, so that control variables are not needed to help identify what causes what.

43. Suppose that you were asked to conduct a study to determine whether male ISU employees are unfairly paid more than female ISU employees are paid. For each ISU employee, you have data on the employee's salary and their gender. Using these salary data as the outcome, would you need control variables for this study?

A. Yes
B. No

Determining whether male ISU employees are *unfairly* paid more than female ISU employees are paid is a causal question, so that control variables will be useful to help identify what causes what, such as whether male ISU employees are paid more than female ISU employees are paid merely because male ISU employees work more hours than female ISU employees work.

44. Suppose that, at a certain university, all students must complete an internship. Some students choose to do the internship during the summer, and some students choose to do the internship during the fall or spring. Suppose that we wanted to assess whether completing the internship during the summer causes a student to be more likely or less likely to get a job within a year of graduation, compared to completing the internship during the fall or spring. One problem for our analysis is that the type of student who completes an internship during the summer might differ from the type of student who completes an internship during the fall or spring, and that difference might affect the likelihood that the student gets a job within a year of graduation. Indicate a control variable that can help address this concern, and explain why that control variable can help address this concern.

Lots of potential acceptable responses, such as the GPA of the student, which might be correlated with selection of the internship time (summer or during the fall or spring) and plausibly influences the likelihood that the student gets a job within six months of graduation.

45. Suppose that you are assigned to test the hypothesis that a female candidate having long hair (compared to having short hair) will affect the percentage of votes that the female candidate receives. For a large set of female candidates, you have data on the percentage of votes that the female candidate received in their election and data on whether that female candidate had long hair or short hair in that election. [1] Identify a control variable that you could use for the analysis. [2] Then explain how that particular control variable would help improve the analysis.

There are many acceptable responses for this item. One control variable would be the political experience of the female candidate. It is plausible that, in general, a candidate with more political experience should be more likely to win an election (such as a state senator being more likely to be elected to the U.S. senate, compare to a candidate that has no political experience). Therefore, if female candidates who had shorter hair tended to have different levels of political experience than female candidates with longer hair, then any association that we see between hair length and winning an election might be due to the political experience and not the hair

length; if we control for hair length, then we have removed hair length as a potential explanation.

[Items 46 and 47] Below are data on the salary and experience of four male teachers and three female teachers.

Teacher	Gender	Salary	Experience	Teacher	Gender	Salary	Experience
1	Male	50	Low	5	Female	50	Low
2	Male	90	High	6	Female	50	Low
3	Male	90	High	7	Female	90	High
4	Male	90	High				

46. Calculate the gender gap in mean salary, to the nearest dollar, between the mean salary of the four male teachers and the mean salary of the three female teachers.

Male teachers: **(50+90+90+90)/4** **=** **80**
Female teachers: **(50+50+90)/3** **=** **63.3**

Gender gap in mean salary is $17.

47. Compared to the gender gap in mean salary when not controlling for experience, indicate whether the gender gap in mean salary when controlling for experience would be larger, smaller, or the same size.

Smaller. The gender gap controlling foe experience should be zero, because experience completely explains salary: each teacher with high experience has a salary of 80, and each teacher with low experience has a salary of 50, so that the gender of the teacher does not seem to matter.

[Items 48 and 49] The table below contains information about six police officers.

Officer	Does the Officer Wear a Body Camera?	Officer Gender	Citizen Complaints about the Officer	Officer	Does the Officer Wear a Body Camera?	Officer Gender	Citizen Complaints about the Officer
1	Yes	Female	6	4	No	Female	12
2	Yes	Female	6	5	No	Male	18
3	Yes	Male	12	6	No	Male	18

48. Based on the table data and to one decimal place, what is the average gap in the mean number of citizen complaints about police officers, between police officers who wore a body camera and police officers who did not wear a body camera?

police officers who wore a body camera: (6+6+12)/3 = 8.0
police officers who did not wear a body camera: (12+18+18)/3 = 16.0

police officers who wore a body camera had 8.0 fewer complaints, on average

49. Based on the table data, statistical control for the gender of the police officer makes the body cameras seem ___ at reducing citizen complaints about an officer, compared to an analysis of the data without any statistical control.

A. less effective
B. as effective
C. more effective

Let's compare female officers:

Female officer + Body camera = mean number of complaints of 6
Female officer + No body camera = mean number of complaints of 12

Let's compare male officers:

Male officer + Body camera = mean number of complaints of 12
Male officer + No body camera = mean number of complaints of 18

The female officer gap is 6, and the male officer gap is 6, so the overall gap when controlling for gender of the officer is 6 complaints on average. So the uncontrolled analysis suggests that the body camera reduce complaints by an average of 8 fewer complaints per officers; but with statistical control for gender of the officer, the body cameras seem to reduce complaints by only 6 fewer complaints per officer, so that statistical control makes the body cameras seem less effective at reducing citizen complaints. The pattern for this is that female officers get fewer complaints and female officers are more likely to wear a body camera, so that imbalance makes the body cameras look more effective at reducing complaints than the body cameras really are.

[Items 50 and 51] The table below contains information about six police officers.

Officer	Does the Officer Wear a Body Camera?	Officer Age	Citizen Complaints about the Officer
1	Yes	20	20
2	Yes	20	20
3	Yes	40	10
4	No	20	30
5	No	40	20
6	No	40	20

50. Based on the table data and to one decimal place, what is the average gap in the mean number of citizen complaints about police officers, between police officers who wore a body camera and police officers who did not wear a body camera?

police officers who wore a body camera: (20+20+10)/3 = 16.7
police officers who did not wear a body camera: (30+20+20)/3 = 23.3

police officers who wore a body camera had 6.6 fewer complaints, on average

51. Based on the table data, statistical control for the age of the police officer makes the body cameras seem ___ at reducing citizen complaints about an officer, compared to an analysis of the data without any statistical control.

A. less effective
B. as effective
C. more effective

Let's compare younger officers:

Younger officer + Body camera = mean number of complaints of 20
Younger officer + No body camera = mean number of complaints of 30

Let's compare older officers:

Older officer + Body camera = mean number of complaints of 10
Older officer + No body camera = mean number of complaints of 20

The younger officer gap is 10, and the older officer gap is 10, so the overall gap when controlling for age of the officer is 10 complaints on average. The pattern for this is that older officers get fewer complaints and older officers are less likely to wear a body camera, so that imbalance makes the body cameras look less effective at reducing complaints than the body cameras really are.

---

52. Data below are from six employees of a company.

Employee	Gender	Pay	Hours Worked
1	Male	$40	40
2	Male	$40	40
3	Male	$20	20
4	Female	$30	40
5	Female	$10	20
6	Female	$10	20

Based on these data and to the nearest penny, what is the gender gap in mean pay among these six employees when controlling for the number of hours worked?

A. $0.00
B. $10.00

C. $16.67
D. $20.00

Among workers who work 40 hours, males make $40 on average and females make $30 on average, so males make $10 more on average than females. Among workers who work 20 hours, males make $20 on average and females make $10 on average, so males make $10 more on average than females. So, across each numbers of hours worked, males make $10 more on average than females.

53. Data below are from eight employees of a different company.

Employee	Gender	Pay	Hours Worked	Employee	Gender	Pay	Hours Worked
1	Male	$30	40	5	Female	$30	40
2	Male	$10	20	6	Female	$10	40
3	Male	$10	20	7	Female	$10	20
4	Male	$10	20	8	Female	$10	20

Based on these data, with no statistical control, the mean pay is $15 among the four male employees and is $15 among the four female employees. Controlling for only the number of hours worked would suggest that ___.
A. there is no gender gap in mean pay
B. there is a gender gap in mean pay that disadvantages male employees on average
C. there is a gender gap in mean pay that disadvantages female employees on average

Males are paid just as much as females, but females work more hours on average, so – with a control for hours worked – females are underpaid.

8.2 Multiple linear regression

[Items 54 and 55] Suppose that researchers are interested in testing whether respondent marital status influences their ratings about the #MeeToo movement. The first linear regression below uses data from the ANES 2020 Time Series Study to predict respondent ratings about the #MeToo movement (FTMETOO), using a predictor for respondent marital status. The second linear regression adds a predictor for respondent age.

```
. reg FTMETOO i.MARITAL
------------------------------------------------------------------------------
      FTMETOO |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
  (intercept) |       55.8        0.5   108.32   0.000         54.8        56.8
              |
      MARITAL |
      Widowed |        1.8        1.6     1.12   0.264         -1.3         4.9
     Divorced |        4.7        1.1     4.24   0.000          2.5         6.9
    Separated |        5.2        2.9     1.78   0.075         -0.5        11.0
Never married |        9.9        0.9    10.65   0.000          8.1        11.8
------------------------------------------------------------------------------

. reg FTMEETOO i.MARITAL AGE
```

FTMETOO	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(intercept)	64.2	1.5	42.20	0.000	61.2	67.2
MARITAL						
Widowed	4.1	1.7	2.44	0.015	0.8	7.5
Divorced	5.1	1.1	4.48	0.000	2.9	7.3
Separated	5.0	3.0	1.67	0.096	-0.9	10.8
Never married	7.3	1.0	7.01	0.000	5.2	9.3
AGE	-0.1	0.0	-5.61	0.000	-0.2	-0.1

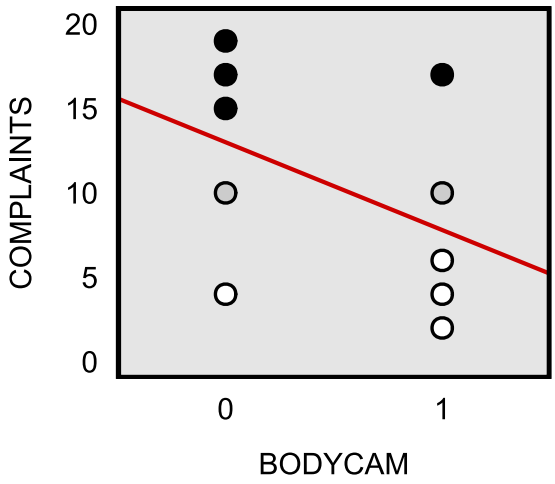
54. Which of the following better indicates the correct interpretation of the 4.1 coefficient for the "Widowed" category in the second regression?
    A. The predicted difference in FTMETOO between a widowed respondent and a married respondent.
    B. The predicted difference in FTMETOO between a widowed respondent and a married respondent, controlling for respondent age.

55. Explain a benefit of the second regression controlling for respondent age.

    Respondents who have different marital statuses might have different ages on average; for example, in the ANES 2020 Time Series Study, the average age for married respondents was 54, but the average age for widowed respondents was 71. If we want to better isolate the effect of marital status, we can compare respondents with different marital statuses at the same age. So, for example, instead of comparing the average married respondent (age 54) to the average widowed respondent (age 71), we would compare a married respondent at age 54 to a widowed respondent at age 54, and then compare a married respondent at age 55 to a widowed respondent at age 55, and so forth, and then average these comparisons.

8.3 Illustration of the effects of statistical control

56. The plot below reports on the x-axis whether the police officer wears a body camera (0 for No, and 1 for Yes) and on the y-axis the number of complaints received by a police officer. The color of the points indicates the officer's age, with black being age 20, gray being age 30, and white being age 40. The slope of the line in the plot is the association between X and Y without any controls, suggesting that the body camera associates with 5.2 fewer complaints on average.



If we predict complaints using the bodycam variable and a control for the age of the officer, that analysis will suggest that the body camera associates with ___.

A. 5.2 fewer complaints on average
B. less than 5.2 fewer complaints on average
C. more than 5.2 fewer complaints on average

For each age (colored black, gray, and white), the line of best fit is flat. This flat line suggests that – controlling for age – the body cam did not affect the number of complaints.