

Name \_\_\_\_\_

## POL 138 Quantitative Reasoning in Political Science Practice · Most missed on Spring 2025 Exam 1

[Items 1 and 2] Suppose that Amy randomly samples 100 ISU students, Bob randomly samples 200 ISU students, and Amy and Bob each measure the height of each student in their samples.

1. Amy calculates a 95% confidence interval for the mean height in her sample, and Bob calculates a 95% confidence interval for the mean height in his sample. Whose 95% confidence interval is more likely to be wider?  
 Amy's 95% confidence interval  
 Bob's 95% confidence interval  
 There is no reason to expect Amy's 95% confidence interval to be wider or thinner than Bob's 95% confidence interval
2. Which statement below is correct?  
 Compared to the mean height in Amy's sample, the mean height in Bob's sample will certainly be closer to the true mean height among the population of ISU students.  
 Compared to the mean height in Amy's sample, the mean height in Bob's sample will probably be closer to the true mean height among the population of ISU students.  
 There is no reason to expect either of the above.

**Commented [LZ1]:** Amy has less data, so her estimate will be less certain. Less certain is wider, in terms of confidence intervals, because Amy has less certainty about the true mean height.

**Commented [LZ2]:** Bob's larger sample should get him closer to the true mean height, but – due to random error – this is not guaranteed.

- 
3. If Hispanics are 19 percent of a sample and 8 percent of the population, what weight should be applied to each Hispanic person in the sample, if weighting on only race/ethnicity?

$8 \div 19$

**Commented [LZ3]:** Survey weights are calculated as population percentage divided by sample percentage.

4. The output below is from a linear regression on data from the ANES 2020 Time Series Study that used a participant's education (in a variable called **EDUC**) to predict the participant's rating about scientists (in a variable called **FTSCIENTISTS**). The **EDUC** predictor is coded so that 0 is a high school degree only, 1 is some college, 2 is a BA/BS 4-year college degree only, and 3 is a graduate degree.

FTSCIENTISTS	Coef.
EDUC	
Some college	3.15
BA/BS degree	6.70
Grad degree	9.90
Intercept	74.56

What does the 3.15 coefficient estimate for "Some college" indicate?

**Compared to participants who had a high school degree only, participants who had some college were predicted to rate scientists 3.15 units higher.**

**Commented [LZ4]:** For a linear regression, coefficients for an included category are compared to the omitted category.

[Items 5 and 6] Suppose that the Illinois Department of Revenue randomly selects 5 percent of Illinois residents for a tax audit. We have data on the county of residence for each of these Illinois residents selected for a tax audit, and, for each county, we calculate the percentage of county residents whom the Illinois Department of Revenue selected for a tax audit. Due to random chance, the percentage residents selected for a tax audit is not exactly 5 percent in each county: the lowest observed percentage of residents selected for a tax audit is 3 percent in one of the counties, and highest observed percentage of residents selected for a tax audit is 7 percent in another of the counties.

5. Which of the following counties is more likely to have been the county in which the observed percentage of county residents selected for a tax audit is 3 percent?
- Cook County, which has 5,200,000 residents
  - Hardin County, which has 3,550 residents
6. Which of the following counties is more likely to have been the county in which the observed percentage of county residents selected for a tax audit is 7 percent?
- Cook County, which has 5,200,000 residents
  - Hardin County, which has 3,550 residents

---

7. Suppose that, in a POL 138 class, the mean final exam score is 90 among students who attended each class meeting and is 80 among students who did not attend each class meeting. Suppose that the p-value is  $p=0.01$  for a test of the null hypothesis that the mean final exam score among students who attended each class meeting equals the mean final exam score among students who did not attend each class meeting. Does this provide sufficient evidence to conclude, at the conventional level in political science, that attending each class meeting caused students to get a higher final exam score, at least on average in that POL 138 class?
- Yes
  - No

8. Suppose that, in Amy's sample, the 95% confidence interval for the mean political ideology is [30, 40]. Which of the intervals below is more likely to be the 99% confidence interval for the mean political ideology in Amy's sample?
- [27, 43]
  - [33, 37]

9. Which of these is closest to what an inference is?
- a conclusion
  - a flawed idea
  - a reason for a prediction

10. A researcher tested the null hypothesis that an association is zero. The p-value for this test is  $p=0.90$ . Based on this p-value, a researcher using the conventional level in political science should \_\_\_\_.
- conclude that the association is zero
  - conclude that the association is not zero
  - do neither of the above

11. 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?
- Yes
  - No

**Commented [LZ5]:** Generally speaking, compared to larger samples, smaller samples are more likely to be farther from the true value. So, compared to Cook County, Hardin County is more likely to be have an outlier low percentage or an outlier high percentage.

This is like comparing Person A who flips a coin one time to Person B who flips a coin 100 times. Compared to Person B, Person A is more likely to get 0% heads and is more likely to get 100% heads.

**Commented [LZ6]:** Students were not randomly assigned to attend each class meeting, so there are alternate explanations to consider. Maybe, for example, the students who attended each class meeting studied more for the exam, so that this studying and not class attendance caused the difference in scores.

**Commented [LZ7]:** To be more certain (99% compared to 95%) we must make a wider range of guesses at the true value.

**Commented [LZ8]:** "Statistical significance" merely refers to sufficient evidence against the null hypothesis.