

GOVT 6019: Midterm Practice Questions

Q1

A research assistant sends you a memo with five claims. Each one contains a mistake. For each claim, (i) name the problem and (ii) explain in one or two sentences what is wrong or what the analyst should do instead.

- a) “We measure protest activity using reports from one national newspaper, which rarely covers rural areas. We recorded 3,000 events, so with this much data any undercount of rural protests will average out.”
- b) “States with more union members have higher minimum wages. This shows that union members support higher minimum wages.”
- c) “To get national turnout, we averaged the 50 states’ turnout rates and got 61.3 percent.”
- d) “Among incumbents who lost re-election, 80 percent had been involved in a scandal. So an incumbent involved in a scandal has an 80 percent chance of losing.”
- e) “Our 95 percent confidence interval for the share of voters who support the measure is [.44, .50]. So there is a 95 percent probability that true support lies between 44 and 50 percent.”

Q2

Every bill in a state legislature is referred to exactly one of three committees. Half of all bills go to Finance, 30 percent to Health and 20 percent to Education. Finance reports out (passes) 40 percent of its bills, Health 60 percent and Education 50 percent.

- a) What share of all bills pass committee? Show your work.
- b) Compute $P(\text{Finance} \mid \text{Pass})$ and $P(\text{Pass} \mid \text{Finance})$. In words, what does each measure?
- c) Are committee and passing independent? Use one numerical comparison.
- d) A reporter writes: “Finance produces more passed bills than any other committee, so it is the easiest committee to get a bill through.” Evaluate this claim using your answers above.

Q3

A candidate holds two town halls in different cities. At each, protesters show up with probability .3, independently of the other. Let X be the number of town halls with protesters.

- a) Write out the PMF of X .
- b) Check that your PMF is valid, and compute $P(X \leq 1)$.
- c) Compute $V[X]$. (Hint: $V[X] = E[X^2] - (E[X])^2$.)
- d) Security costs \$5,000 for each town hall with protesters, plus a fixed \$2,000 for the campaign’s security plan. Find the variance and the standard deviation of the security cost. You may leave square roots unevaluated.
- e) Suppose instead both town halls are in the same city on the same weekend. Is independence still plausible? How would $V[X]$ change, and why?

Q4

In County X, 2 percent of in-person provisional ballots and 10 percent of mail ballots were rejected. The county's overall rejection rate across both types was 4 percent.

- What share of County X's ballots were mail ballots? Show your work.
- County Y rejected 1 percent of in-person ballots and 8 percent of mail ballots. Half of its ballots were mail. Compute County Y's overall rejection rate.
- County Y has a lower rejection rate than County X for *each* ballot type, yet a higher overall rate. Explain how.
- County X counted 9,000 ballots and County Y 1,000. A reporter averages the two overall rates and reports a combined rate of 4.25 percent. What share of all 10,000 ballots was rejected? Why does it differ from the reporter's number?
- A third county, Z, reported 0 ballots returned and -99 rejected, where the codebook says -99 means "data not available." What does this code print for County Z, and what should the analyst have done first?

```
d <- tibble(county = c("X", "Y", "Z"),
             returned = c(9000, 1000, 0),
             rejected = c(360, 45, -99))
d |> mutate(rate = rejected / returned) |> pull(rate)
```

Q5

A simple random sample of 100 adults in a city records how many minutes of local news each watched yesterday. The sample mean is 22 minutes. The researcher reports a 95 percent confidence interval of 22 ± 3.9 , or about $[18.1, 25.9]$, using the multiplier 1.96.

```
> 3.9 / 1.96
[1] 1.989796
> 2 * pnorm(-1)
[1] 0.3173
```

- What is the estimate of the population mean, and what is the estimand?
- Find the standard error. Round to a whole number.
- Test $H_0 : \mu = 20$ against a two-sided alternative at $\alpha = .05$. Give the test statistic and your conclusion.
- Interpret the p -value from (c).
- About how many adults would the researcher need to cut the interval's width in half? Explain.
- A city council member says: "So 95 percent of residents watch between 18 and 26 minutes of local news a day." What is wrong? Use the standard error to estimate the standard deviation of viewing time across individuals.
- Viewing time is strongly right-skewed: most people watch little and a few watch a lot. Does this make the confidence interval invalid? Explain.

- h) If 200 researchers each drew their own sample of 100 and built a 95 percent interval this way, about how many intervals would fail to contain μ ?

To estimate how much viewing time varies across residents, the researcher considers two estimators of the population variance σ^2 :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

You may use the facts that $E[S^2] = \sigma^2$ and $E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2$.

- i) Which estimator is unbiased? Is the other one too high or too low on average, and by what fraction when $n = 10$ and when $n = 100$?
- j) Is $\hat{\sigma}^2$ consistent? Explain what happens to its bias as n grows.
- k) In one or two sentences, give the intuition for why dividing by n underestimates σ^2 .