

Quantitative Political Methods

Midterm 1 Review

Be able to define these terms:

Describing data:

1. Statistics
2. Data
3. Observations
4. Population
5. Sample
6. Descriptive vs. inferential statistics
7. Parameter vs. statistic
8. Variable
9. Qualitative/quantitative data
10. Nominal/ordinal/interval
11. Granularity (continuous/discrete)
12. Skew (positive & negative)
13. Outliers

Measures of Central Tendency and Dispersion:

14. Mean
15. Median
16. Range
17. Variance
18. Standard deviation
19. IQR

Sampling:

20. Simple random sample
21. Systematic random sample
22. Stratified random sample
23. Cluster sample
24. Multi-stage sampling
25. Sampling bias
26. Sampling error
27. Selection bias

28. Response bias
29. Social desirability effect
30. Order effect
31. Non-response bias

Distributions

32. Frequency distribution
33. Binomial Distribution
34. Normal Distribution
35. T Distribution
36. Degrees of Freedom
37. Probability density
38. Standard deviation/variance
39. Empirical rule
40. Sampling distribution
41. Central Limit Theorem
42. Standard error

Estimation

43. Point estimate
44. Interval estimate
45. Confidence interval
46. Confidence level
47. Bias
48. Efficiency
49. Margin of error

Hypothesis testing

50. Hypothesis
51. Null/alternative hypotheses
52. Test statistic
53. P-value
54. Significance level (α -level)
55. Type I & Type II error

Be able to calculate/execute these concepts:

Describing Data:

1. Classify variables as nominal/ordinal/interval
2. Classify variables as continuous/discrete

Measures of position:

3. Make a stem and leaf plot
4. Find percentiles/quantiles
5. Find IQR and outliers
6. Create a boxplot
7. Identify the skew of a data set

Measures of central tendency:

8. Mean
9. Median
10. Mode

Measures of dispersion:

11. Range
12. Variance
13. Standard deviation
14. IQR

Distributions and probability:

15. Define a sampling distribution of a mean
16. Define a sampling distribution of a proportion
17. Calculate a Z-score
18. Use the Z-table (find proportions given a Z-score or a Z-score given proportions)
19. Find probabilities using the Normal distribution
20. Use the T-table
21. Find probabilities using the T-distribution
22. Find probabilities using the binomial distribution table

Confidence intervals:

23. Confidence interval for a mean (large sample)
24. Confidence interval for a mean (small sample)

One-Sample Hypothesis Testing:

25. Hypothesis testing for a mean (large sample)
26. Hypothesis testing for a mean (small sample)
27. Hypothesis testing for a proportion (large sample)

Practice problem set:

1. Identify each variable as nominal/ordinal/interval and discrete/continuous:
 - a. Type of car driven (nominal, discrete)
 - b. General health (poor, reasonably good, excellent) (ordinal, discrete)
 - c. College tuition (interval, continuous)
 - d. Number of political parties in a country
 - e. Religious affiliation
 - f. Distance between home and work
2. Create a boxplot for the following dataset: 4, 8, 32, 41, 16, 11, 1, 38, 3, 27, 48, 19
Is this data skewed? If so, in which direction?

3. The following table contains the GDP per capita (in thousands of international dollars) for four European countries.

Belgium	Germany	France	Luxembourg
38	38	35	90

- Find the mean.
- Find the median.
- Find the mode.
- Find the range.
- Find the variance.
- Find the standard deviation.
- Would you say that one of these observations is an outlier?

4. The distribution of math SAT scores for students that get into Washington University in St. Louis follows a normal distribution with a mean of 740 and a standard deviation of 23. If we poll a random sample of 51 students, what is the probability that the mean score of your sample will be 746 or below?

5. The Freshman Fifteen is an expression that commonly refers to an amount (somewhat arbitrarily set at fifteen pounds) of weight often gained during a student's first year at college. You decide to test whether this expression holds true for Washington University in St. Louis. You randomly select 16 sophomores and gather data on how much weight (in pounds) they gained the previous year. The mean of your data is 14.5 lbs. and the sample standard deviation is 0.8 lbs.
 - a. Identify the population for this study.

 - b. Describe the sample distribution for this study.

 - c. Describe the sampling distribution for this study as precisely as possible.

- d. Calculate the point estimate and a 95% confidence interval for the population mean. Explain what your confidence interval means.

6. Apple claimed that iPhone 5 is “the biggest thing to happen to iPhone since iPhone.” Among other improvements, the iPhone 5 claimed improved battery life over the old versions. For example, the standby time has been improved to 225 hours (a 25-hour improvement over iPhone 4S). To test this claim, you collect a sample of battery longevity from 100 randomly selected owners of the iPhone 5. Among these 100 owners, you find that the battery life in the new iPhone 5 is 217 hours with a standard deviation of 40 hours. Test the research hypothesis that *the batteries in the new iPhone 5 differ from the 225 hours claimed by Apple*. Use a 0.05 level of significance.

7. You roll a 6-sided die.
 - a. What's the probability of rolling a four three times in a row?
 - b. Assume I roll the dice four times and get the same number each time. I am suspicious and suspect that the die is unfair. Test my hypothesis using a 0.05 level of significance.
8. Imagine you are interested in the different patterns of support for the Spanish government among citizens of Catalan population. You decide to conduct a survey asking people "Do you have confidence in the national government?" Possible answers include Yes or No. You were able to poll 243 Catalans. Of these 243 respondents, 86 said, "Yes."
 - a. Provide a point estimate for the percent of Catalans that have confidence in the government.
 - b. Identify the sampling distribution of this study. Be precise.

c. Construct a 92% confidence interval of the percent of Catalans that have confidence in the government.

d. Test the theory that less than 40% of Greeks support the government using a 0.05 significance level.

9. Suppose a random sample is taken of 200 Jedi Knights. The mean number of droids killed by a Jedi Knight is 19, with a standard deviation of 2. Construct and interpret a 92% confidence interval for the mean number of droids killed.
10. The distribution of Quantitative GRE scores for graduate students who drop out of graduate school has a mean of 550 and standard deviation of 30. These scores are distributed normally.
- If I select one graduate school dropout at random, what is the probability their score was 574 or below? (You can assume that the population is normally distributed)
 - If you take a sample of 100 graduate school dropouts, what is the probability that the average Quantitative GRE score of your sample is 574 or below? Explain why this number is different from the answer you put down for part (a).

- c. Assume that you do **not** know the true population mean and standard deviation. You take a random sample of 24 graduate school dropouts and calculate a mean of 570 and a standard deviation of 16. Construct a 95% interval for the true population mean.
- d. The average GRE scores for all graduate students is 582.5. Using only the sample described in part (c), test the research hypothesis that graduate school dropouts have lower GRE scores than average. Specify all five steps for conducting a hypothesis test. Choose your own level of significance.