

## Practice Questions for Exam

1. Identify each of the following as either qualitative or quantitative and either discrete or continuous. If the variable is quantitative, does it have an interval or ordinal scale?
  - (a) The number of bedrooms in your apartment
  - (b) The GDP of a country.
  - (c) The type of political system of a country (e.g. Democracy, Aristocracy, etc.).
  - (d) Access to clean water (yes, no).
  - (e) Public expenditures in education.
  - (f) Numbers of days a week you read the newspaper.
  - (g) Party affiliation.
  
2. You are interested in the relationship between a person's party affiliation and her parents' party affiliation. You have collected this data for students in our class. How could you describe the relationship using a graphical or tabular technique?

3. You have collected the income of students in our class. How could you summarize the distribution of this variable numerically?
4. Suppose that the correlation between a countries GDP and national debt of these countries is -0.8. How would you describe this relationship?
5. According to a recent Washington Post / ABC News poll of 1,000 adults, 11% think that the number of U.S. military forces in Iraq should be increased. Form a 95% confidence interval for this proportion.
6. The 2004 Latinobarometer asked respondents “how fair is income distribution in your country?” Table 1 shows the results for 969 respondents in Brazil

How fair is income distribution?	Frequency
Very fair	10
Fair	126
Unfair	514
Very unfair	319
Total	969

- (a) How could you describe these data graphically?
  - (b) Report the median and the mode. Interpret.
7. A survey group asked respondents to place themselves on a left-right political ideological scale going form 0 (meaning left) to 10 (meaning right). The results for Mexico and Colombia are displayed in Table 2.

Ideological Scale	Percentages	
	Mexico (N= 927 )	Colombia (N= 716)
0 Left	6 %	2 %
1	3 %	4 %

2	5 %	2 %
3	8 %	3 %
4	13 %	2 %
5	35 %	26 %
6	9 %	6 %
7	7 %	6 %
8	6 %	8 %
9	3 %	8 %
10 Right	5 %	34 %

- Construct a 95% confidence interval around the difference in reported proportions for the two extreme categories, 0 and 10. Interpret.
8. A car manufacturer claims that its cars will run for an average of 20,000 miles before needing their first repair. To prove its claim, it has tracked a random sample of 21 cars. It found that the sample average number of miles before repair was 18,700, with a standard deviation of 8,600 miles. Test the null hypothesis that the population mean is 20,000 miles against the alternative that it is not.
  9. A person claiming to possess extrasensory perception (ESP) says she can guess more often than not the outcome of a flip of a balanced coin in another room, not visible to her. Of the ten coin flips, she guesses the correct result seven times. Does she possess ESP?
  10. A deli you frequent claims that their sandwiches contain  $\frac{1}{4}$  of a pound of meat. You are quite suspicious of this claim and have measured and recorded the amount of meat in the last 17 sandwiches you purchased. You determine that the average amount of meat was 0.23 pounds with a standard deviation of 0.056 pounds. Are you being cheated?
  11. A group of female employees are suing Wal-Mart claiming gender discrimination. Out of the 27 employees in their store, there are 11 females and 16 males. The

- males had an average salary of \$14 per hour with a standard deviation of \$3 and the females had an average salary of \$11 per hour with a standard deviation of \$1. Is there any merit to their claim?
12. In a survey conducted in 1992, senior high school students were asked if they had ever used marijuana. Of the females sampled, 445 said yes and 675 said no; of the males sampled, 515 said yes and 641 said no. Are male high school students more likely to use marijuana?
13. A random sample of 834 homes in Alabama and 717 houses in Mississippi were collected. The homes in Alabama had an average value of 114 thousand with a standard deviation of 13. The homes in Mississippi had an average value of 121 with a standard deviation of 18. Are real estate prices higher in Mississippi?
14. Consider the SPSS data file 'house.sav' available on the course website.
- (a) Is the distribution of district population (`cd_pop`) unimodal, bimodal, or neither? Justify your answer with a graphical summary.
- (b) The variables "dem\_cand\_id" and "rep\_cand\_id" are measures of the ideological positions of Democratic and Republicans House candidates in each congressional district. What can you say about the relationship between these two variables? Justify your answer with a numerical measure.
- (c) Test the null hypothesis that southern and non-southern congressional districts are equally conservative, using the 1% significance level. Interpret the result of this test. Use the variable 'id\_survey' as the measure of district ideology. The variable "cd\_south" is equal to 1 if the congressional district is in the south and 0 otherwise.

15. Consider the data file 'stocks.sav'. Produce a line graph representing the value of S&P 500 index over time. Use the variables 'SPClose' and 'Date'. Interpret the results of the plot.
16. Consider the data file 'ises01.sav', which is a survey of Israelis after the 2001 prime-ministerial election. What is the relationship between a respondent's ideology and their attitude about returning the Golan Heights to Syria? Use the variable A66 (Right left) as your measure of respondent ideology and A22 (Return Golan to Syria) as your measure of support for returning the Golan Heights to Syria. Do the results conform to your expectations?
17. You have been hired as a consultant to advise a candidate on polling methodology. Consider the following weighting variables? Which would you recommend that the candidate use in adjusting their sample (or does it depend)? Explain.
  - (a) Income.
  - (b) Education level.
  - (c) Partisan identification.
  - (d) Vote in the previous presidential election.
18. Suppose that you use the Gallup method to construct an index for likely voting which ranges from 0 to 7. You find that 40% of respondents have a score of 7, 20% have a score of 6, 15% have a score of 5, 10% have a score of 4, 10% have a score of 3, 1% have a score of 2, 1% have a score of 1, and 3% have a score of 0.
  - (a) Suppose that you to make a projection based on 50% turnout. What percentage of each of the groups (scores 0-7) will be included in your sample?

- (b) What if you want to make a projection based on 75% turnout. What percentage of each of the groups (scores 0-7) will be included in your sample?
19. A journalist makes the following statement: “I don’t trust the Gallup poll. There are simply not enough educated people in their sample, and consequently, I believe the poll will underestimate the proportion of voters supporting Democratic candidates”. Do you agree or disagree with this statement? Explain and give the assumptions that your answer depends on.
20. A journalist makes the following claim. “I don’t trust the LA Times poll. The sample contained only 27% Republicans, while we know from the last exit poll that 35% of voters identify with the Republican Party”. Do you agree or disagree with this statement? Explain and give the assumptions that your answer depends on.
21. Consider the data file wp04.sav on the course web page. Consider the item q19 which measures the respondent’s evaluation of whether the war in Iraq was worth fighting.
- (a) What is the sample proportion of respondent’s who answer “Yes, worth fighting, STRONGLY”?
- (b) What is the weighted sample proportion of respondent’s who answer “Yes, worth fighting, STRONGLY”?
- (c) Provide one reason why these two quantities may differ.
- (d) Which provides a better estimate of the population proportion who would answer, “Yes, worth fighting, STRONGLY”? Explain.

22. Consider the data file `gallup06.sav` on the course web page. This is the final Gallup pre-election poll before the 2006 Congressional elections. The data file contains both a demographic weight (`wtfctr`) and a series of likely voter weights (`wtlv35`, `wtlv40`, `wtlv45`, `wtlv50`).
- (a) Report the predicted percent of voters who will vote for a Democratic congressional candidate (`q2`) using both the demographic weight “`wtfctr`” and the likely voter weights (`wtlv35`, `wtlv40`, `wtlv45`, `wtlv50`).
- (b) Why do these estimators differ?
23. Suppose you were hired by Christine O’Donnell’s Senate Campaign immediately after she won the Republican nomination in Delaware. The campaign would like you to conduct a poll and analyze the results. The campaign would like to avoid the impression that she has no chance on winning on Election Day by releasing poll numbers that suggest that the race is closer than the other polls are suggesting. You are instructed to make the results look as good for Christine O’Donnell as possible. What methodological choices (short of making up data) could you make that would have the likely effect of inflating her poll numbers?
24. Do you agree or disagree with the following statement: “I don’t trust telephone surveys anymore. More and more individuals—particularly young individuals—do not have a land line. Moreover, these individuals are likely to differ from older individuals on key issues such as support for Democratic candidates, acceptance of homosexuality, and support for the legalization of marijuana. If we are missing these younger individuals, our survey estimates will be biased”. Explain.