

**DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Posc/Uapp 815**

SAMPLE QUESTIONS FOR THE FINAL EXAMINATION

I. READING:

- A. Read Agresti and Finalay, Chapters 6, 7, and 8 carefully.
1. Ignore the sections on testing proportions, although they simply apply the methods to percentages.
 2. You should know what a type one error is and the probability of making one in a standard hypothesis test situation.
 3. You should understand confidence intervals and their connection to hypothesis tests.
 - i. See pages 177 to 178. The first paragraph provides a clear explanation. So, for example, if I say the test of a null hypothesis (at the .05 level) that $\mu = 0$ leads to an observed t of 2.67, you should be able to answer the question: would a 95 percent confidence interval include the hypothesized value, 0? (The answer is "no," and the explanation is found on page 177.)
 4. Review all of your notes and especially the ideas about expected values, the mean of sampling distributions, the variance or standard deviation of sampling distributions, and the odds ratios.
 5. You should also know the difference between a mean and median.
 6. We discussed at great length areas under the normal curve.
 7. Be sure that you can read z and t tables and know how to calculate degrees of freedom.
- B. I will ask questions about testing a single mean, a difference of means, and a hypothesis of statistical independence, which is covered in the last class.

II. AREA UNDER THE NORMAL CURVE AND MEANS:

- A. For any normally distributed variable how much of the area under the normal curve lies between the mean (μ) and plus one standard deviation above the mean?
1. That is, how much of the area lies between μ and $\mu + \sigma$?
 2. How much are lies between $\mu - \sigma$ and $\mu + \sigma$?
 3. How much lies between $\mu - 1.96\sigma$ and $\mu - 1.96\sigma$?
 4. How much lies between $\mu - 1.96\sigma$ and $\mu - 2.56\sigma$?
 5. How many standard deviations above and below the mean do you have to go to have 50 percent of the area.
 6. Note: these should be straight forward; just draw pictures if in doubt.
- B. The average amount of gasoline used by American drivers per week is 22 gallons with a standard deviation of 6 gallons.
1. What percent of drivers use more than 28 gallons per week?

2. What percent use less than 16 gallons a week?
 3. You should be able to figure this out in your head if you have table in front of you.
- C. According to the Census Bureau the mean number of years of education for self-employed workers in 1990 is 13.6 years with a standard deviation of 3.0 years.
1. What is the mean and standard deviation of the sampling distribution for the following samples? (Use the formula on page 100.)
 - i. $N = 9$
 - ii. $N = 36$
 - iii. $N = 100$
 2. Sketch the distributions in such a way that it's clear what the effect of sample size on the sampling distribution is.
 - i. Note: we answered this type of question on the first exam.
 3. Suppose each of these samples has a sample mean of $\bar{Y} = 14$. For each sample what is the probability of getting a \bar{Y} that large or larger?
 - i. Note: just calculate a z statistic and look its value up in the table. What proportion of the curve lies above it? That's the probability of getting that \bar{Y} or one larger.
- D. Can You do problem 5.5? The mean of a sample of $N = 637$ subjects number of sex partners is 1.314 with a standard deviation ($\hat{\sigma}$) of 5.418.
1. What is your best estimate of the number of sex partners of the people in this population?
 2. Construct 95 percent confidence intervals for this estimate.
 - i. Note the sample size.
 3. Construct 99 percent confidence intervals.
- E. The previous question was based on 637 subjects. Suppose, however, the researcher managed to interview only 20 people.
1. What are the 95 percent confidence intervals in this case?
 2. What about 99 percent confidence intervals?
 3. Without looking which interval is wider?

III. TEST OF A SINGLE MEAN:

- A. Remember that you have to keep in mind how many cases are involved.
- B. In response to a the statement "A preschool child is likely to suffer if his or her mother works," the possible responses were "Strongly Agree" (coded 2), "Agree" (coded 1), "Disagree" (coded -1), and "Strongly disagree" (coded -2). The question was asked of a simple random sample of 996 Americans. The results are

N	Mean	Std. Deviation	Std. Error of Mean
996	-.052	1.253	.0397

1. An researcher wants to know if the true average (μ) is 0 or not.
 - i. What is the null hypothesis?
 - ii. What is the two-sided null hypothesis?
 - iii. What sampling distribution would you use to help answer the researchers question? Why?
 - iv. What is the observed test statistic?
 - v. What is its approximate P value? (That is, if the null hypothesis is true, how likely–what is the probability–that you would get a statistic as larger or larger than the one you observed?)
 - vi. What is the 95 confidence interval? Does it include the hypothesized value?
 - vii. What would you tell the researcher about public opinion toward women working?
 - viii. By the way, what assumption does one make about the type of variable this question involves?

- C. A school principle, being sensitive to criticisms that her school is under performing compared to national standards asks your advice. According to Department of Education data the national average score on a test 100. She shows you the records for a sample of 20 students in the school. The average of their test scores is 92 with a standard deviation of 5 points. She obviously wants to be able to say that although the sample mean is below the national average, there isn't enough evidence to conclude that the school as a whole is underperforming and should be closed. What can you say about this matter?
 1. Your answer should consist of two parts. The first a formal test of a statistical hypothesis that follows the guidelines discussed in class and the book.
 2. And then a one-paragraph (at most!) explanation that a school administer can understand.

IV. DIFFERENCE OF MEANS TESTS:

- A. Here are statistics on homicides per 100,000 population for two types of counties.

Urban	Rural
$N_1 = 33$	$N_2 = 19$
$\bar{Y}_1 = 57$	$\bar{Y}_2 = 52$
$\hat{\sigma}_1 = 11$	$\hat{\sigma}_2 = 14$
$\hat{\sigma}_{\bar{y}_1 - \bar{y}_2} = 3.57 \quad t_{\text{obs}} = 1.40$	

1. What is the “effect size” estimate? (Note: I mentioned this term as an

- alternative phrase for what?)
2. Test the hypothesis that the population value for this effect is 0.
 3. What is the degrees of freedom for this problem?
 4. Construct 95 percent confidence intervals for the estimated difference or effect size.
- B. Suppose the sample sizes were 330 and 190 respectively but everything else is the same.
1. Answer questions 1 through 4, except one of the questions no longer makes sense. Which one and why?

V. **CROSS CLASSIFICATIONS AND ODDS RATIOS:**

- A. Here is a contingency table that shows the relationship between party identification and gender:

Gender			
	Male	Female	Total
Democrat	165	279	444
Independent	47	73	120
Republican	191	225	416

1. The problem is: to what extent if any do these data suggest the existence of a gender gap in American politics?
2. Think a bit before answering. What are the odds of a female identifying with a party? What about the similar odds of male?
3. How do the odds compare?
4. What are the odds of a female being a Democrat? Of a male? (Exclude “independents” when answering this question.)
5. Are gender and party identification statistically independent?

B.