

POS 3713 Summer 2000
Review Terms and Problems for Midterm Exam #2

The midterm exam is scheduled on Thursday, June 1st. I will hold a review session on Wednesday, May 31st from 5:00-7:00pm in 511 Bellamy.

Review Terms

Descriptive & Inferential Statistics

Variable

Data

Population

Sample (Random sample)

Statistics

Discrete vs. continuous variables

Enumerative table

Frequency

Proportion

Percentage

Contingency (cross-tabulation) table

Mean (characteristics)

Median

Mode

Index of qualitative variation (IQV)

Range

Variance

Standard deviation

Normal curve (characteristics)

Standardized normal curve

Z-scores

Percentile rank

Sampling distribution (importance)

Central Limit Theorem

Point estimates vs. confidence intervals

Bias

Efficiency

Confidence intervals for sample means

Confidence intervals for sample proportions

Hypothesis testing for sample means (large samples, Z and small samples, t)

Hypothesis testing for sample proportions

Critical region

Confidence level

One tailed vs. two-tailed tests of significance

Null hypothesis

Alternative hypothesis

Type I and Type II error

Review Problems

1. The average age for registered voters in Leon county is $\mu = 39.7$ years. During a recent jury trial in the county courthouse, a statistician noted that the average age for the 12 jurors was $X = 50.4$ years with $s = 11.8$. Is it reasonable to conclude (at the 95% confidence level) that this jury is *not* a random sample of registered voters, i.e., that the jury is different from Leon county's population?
2. A random sample of 126 local political science graduates scored an average of 458 on the GRE advanced political science test with a standard deviation of 20. Is this significantly larger than the national average? ($\mu = 440$). Assume 95% confidence, or $\alpha = 0.05$.
3. Consider the following descriptive statistics for the percentage of males and females who read. This data is taken from the 1995 World Survey, and includes data for 85 different countries. Assume that this sample of data in 1995 is drawn from the population of all countries for all years of their existence.

Statistics

| | | Females who read (%) | Males who read (%) |
|----------------|---------|----------------------|--------------------|
| N | Valid | 85 | 85 |
| | Missing | 24 | 24 |
| Mean | | 67.26 | 78.73 |
| Median | | 71.00 | 87.00 |
| Mode | | 100 | 100 |
| Std. Deviation | | 28.61 | 20.45 |
| Variance | | 818.38 | 418.01 |
| Range | | 91 | 72 |
| Minimum | | 9 | 28 |
| Maximum | | 100 | 100 |

- a. Calculate the 95% confidence interval for the mean percentage of females who read; assume this measure is an interval measure (use $N = 85$). How do you interpret this confidence interval?
 - b. Is there a significant difference between the percentage of males and females who can read? Make a comparison based on 1) the measures of central tendency, and 2) the measures of dispersion. What can account for these differences?
4. Soon after he took office in 1963, 160 out of 200 Americans sampled approved of President Johnson. With growing disillusionment over his Vietnam policy, by 1968 he was approved by only 70 out of a sample of 200 Americans. What is the 95% confidence interval for the percentage of all Americans who approved of Johnson in 1963? In 1968? (Hint: use the confidence interval formula for proportions) How do you interpret these confidence intervals?

5. National data on interstate highways reveals that $\frac{3}{5}$ (0.6) of all cars are exceeding the 70-mph speed limit. Governor Bush is interested in maintaining highway safety in Florida, therefore he collects a random sample of 120 cars. This sample from Florida reveals that 84 cars exceeded the posted 70-mph speed limit. Should Governor Bush be concerned about these results? In other words, can he be 95% confident that Florida drivers are *more* likely to speed on interstate highways than drivers in the rest of the country? (Hint: hypothesis test for sample proportion)
6. Calculate the index of qualitative variation (IQV) for party identification using the 1994 National Election Study data presented below. What does the IQV tell you about the dispersion of party identification in the United States?

| <u>Party Affiliation</u> | <u>Frequency</u> |
|--------------------------|------------------|
| Democrat | 612 |
| Independent | 513 |
| Republican | 544 |

6. Using the standard deviations and means produced in the table below, calculate the following for both *ClinTher* and *HillTher* :

Descriptive Statistics

| | N | Range | Minimum | Maximum | Mean | Std. Deviation | Variance |
|-----------------------------|------|-------|---------|---------|-------|----------------|----------|
| Clinton thermometer | 1705 | 100 | 0 | 100 | 59.34 | 29.58 | 874.802 |
| Hillary Clinton Thermometer | 1685 | 100 | 0 | 100 | 52.81 | 29.85 | 891.005 |
| Valid N (listwise) | 1683 | | | | | | |

- a) Probability that the score on the feeling thermometer will be greater than 75; e.g., $\text{prob}(ClinTher > 75) = ?$ and $\text{prob}(HillTher > 75) = ?$ (Note: probability should be expressed as a proportion)
 - b) Probability that the score on the feeling thermometer will be less than 25; e.g., $\text{prob}(ClinTher < 25) = ?$ and $\text{prob}(HillTher < 25) = ?$
 - c) The percentage of people who score Bill or Hillary between 30 and 70 on the feeling thermometer
 - d) The percentile rank for a rating of 85 on the feeling thermometer scale
7. Calculate the mean, median, mode, range, variance, and standard deviation for the following golf scores. Think about how to interpret each of these statistics substantively.

70 72 68 65 71 75 66 67 65 70

8. Refer to the following theoretical hypothesis and contingency table (cross-tabulation) to answer the questions in the space below.

Theoretical Hypothesis: Women are more likely to identify with the Democratic party because it is more sensitive to issues that women care about, such as education, abortion (a pro-choice position), and family issues (such as maternal leave).

Data: Survey data from the 1994 National Election Study (N=1795, but some cases have been dropped for this analysis)

| | Party Identification | | | |
|--------|----------------------|----------------|----------------|----------------|
| Sex | Republican | Independent | Democrat | Total |
| Male | 273 (34.7%) | 269 (34.2%) | 244 (31%) | 786 (47.1%) |
| Female | 271 (30.7%) | 244 (27.6%) | 368 (41.7%) | 883 (52.9%) |
| Total | 544 (32.6%) | 513 (30.7%) | 612 (36.7%) | 1669 (100%) |

1. What is the independent variable for this theory? What is the dependent variable?
2. What *percentage* of the total sample is male?
3. What *proportion* of the total sample is Democratic?
4. What *percentage* of respondents are female Independents?
5. How many females in this sample identify with the Republican party (i.e., what is the frequency)?
6. Does the table provide any evidence to support the theoretical hypothesis? Justify your answer using the information presented.