POL 51: The Scientific Study of Politics
Fall Quarter 2008
Study Guide Final Exam

Multiple Choice (1 point each; write answers in BLUE BOOK)

1. In a distribution of data, if $\overline{X} > median$, the data can be described as being:
a. Normally distributed
b. Right Skewed
c. Unimodal
d. Left Skewed
e. All of the above.

2. The interquartile range is to the median what the _____ is to the mean.
a. normal distribution
b. sampling distribution
c. standard deviation
d. correlation coefficient
e. None of the above.

3. Which is true about the *t*-distribution?
a. It has slightly "fatter" tails than the *z*-distribution.
b. We use the *t* when *s* is used instead of σ for the standard error.
c. As *n* increases, the *t* distribution closely approximates the *z*-distribution.
d. The probability area (density) is a function of the degrees of freedom.
e. All of the above.

4. Under what condition does the median do a better job than the mean, $\overline{X}$, of describing central tendency?
a. When the data are highly skewed.
b. When the data are bimodally distributed.
c. When the data are normally distributed.
d. Never: $\overline{X}$ always is the preferred statistic.
e. None of the above.

5. Which is *not* true about the *z*-distribution?
a. It is symmetrical about the 0 point.
b. It has a standard deviation of 1.
c. Computation of probability area depends on sample size.
d. It is a standardized version of the normal distribution.

6. If the *p*-value for a two-tail test is reported to be .10 ($\alpha = .05$) from statistical software output, the *p*-value for a one-tail test would be?
a. .20
b. .05
c. .10 as well
d. Not enough information given in problem to tell.

7. A 95 percent confidence interval is computed. Imagine that $\mu = 0$. Suppose this interval

does not contain $\mu = 0$. Which of the following is an *incorrect* statement about this confidence interval?
a. "I am 95 percent confident that this confidence interval does not contain $mu = 0$.
b. "In repeated samples, 95 percent of all samples would produce an interval like this one."
c. "In repeated samples, only about 5 percent of all samples would produce an interval containing $\mu = 0$.
d. Each of the above statements are correct.

8. Consider this hypothesis: H_a: Women more favorably evaluate Democratic candidates than when compared to men. Which is true about this hypothesis:
a. It is a normative statement about gender and liberal ideology.
b. It is a non-directional hypothesis.
c. It is a directional hypothesis.
d. None of the above are true.

9. A sampling distribution is best described as:
a. The degree to which the distribution of sample elements are representative of the population.
b. The distribution of statistics from repeated samples of size n from the same population.
c. The number of cases that end up in the final sample.
d. The confidence interval around a statistic.

10. Which is *not* true about one-tailed hypothesis tests?
a. They rely on probability area in either the upper or the lower tail of a distribution.
b. They are less conservative tests than two-tailed tests.
c. For the same α level as a two-tail test, it is harder to reject the null with a one-tail test.
d. They should only be used in conjunction with directional hypotheses.

Scenarios (Please read the following scenarios and answer the subsequent questions):

1. A political consultant is hired to determine the effectiveness of spending on television advertising and unique "hits" to a political website. It is determined in advance that the advertising campaign is only feasible if for every $1,000 spent on the advertising campaign, the average number of *unique* hits to the website per media market is greater than 5,000 (unique means 5,000 different people, not 5 people visiting 1000 times each!). Uniqueness of hits must be estimated and so there will be error around the statistic. To see if this advertising campaign meets expectations, a pilot study is conducted in 10 media markets. The consultant finds the following: for every $1,000 spent in each of these markets, the average number of hits is 5,200 (that is, $\overline{X} = 5200$). However, it is estimated that because of repeat website visitors and other sources of measurement error, the standard deviation around this mean is 345 (that is, s=345). To summarize, then, the mean is 5200, the standard deviation is 345 and the sample size is 10. Please answer the following questions:

a. What is the null hypothesis? (3 points)
b. What is the alternative hypothesis? (3 points)
c. Why did you pose H_a this way? (4 points)
d. Using $\alpha = .025$, what is your decision about rejecting or not rejecting the null. Why/how did you come to this decision? (8 points).
e. If $\alpha = .05$, would your conclusion be different from that stated in d? Why or why not? (5

points)

2. A researcher is interested in comparing the approval ratings of a member of congress from the same congressional district and of the same political party but from different periods of time, say 1986 and 2006. The researcher discovers that the approval rating for the member of congress in this district in 1986 was 37 percent and the approval rating for the member in this district in 2006 was 46 percent. The researcher concludes that the member in 1986 was "far less popular" than the member in 2006. However, another researcher comes along and says, "wait, you need to put these numbers in historical perspective by standardizing them with a *z*-score." She then supplies the following information:

For 1986: $\overline{X}_{Approval} = 44\%$; $s = 10\%$.
For 2006: $\overline{X}_{Approval} = 58\%$; $s = 6\%$.

a. Based on this information, who, in historical context, was "least popular"? Explain precisely why and how you come to this conclusion (10 points).
b. What is the probability of observing an approval rating score as low or lower than that observed for the 1986 member? (2 points)
c. What is the probability of observing an approval rating score as low or lower than that observed for the 2006 member? (2 points)

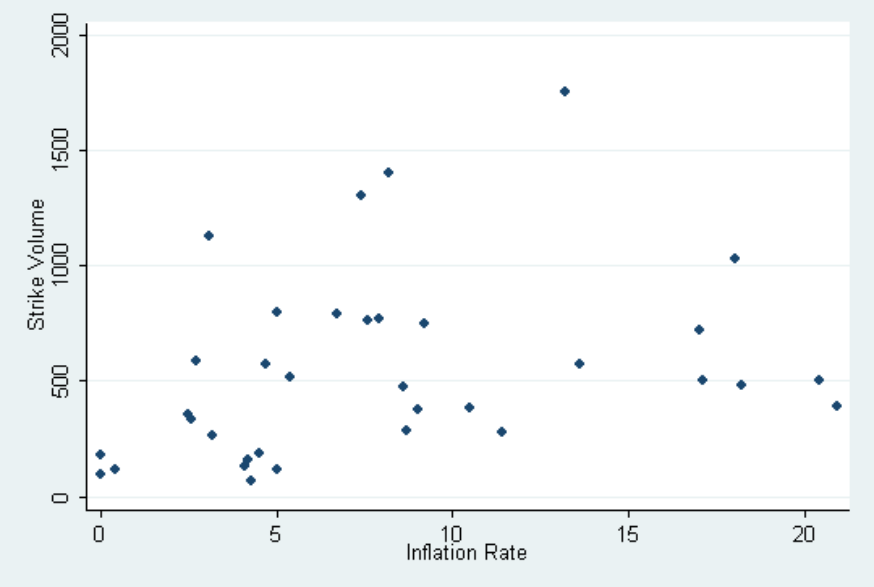
3. A researcher hypothesizes that attitudes toward immigrants will be higher among individuals who reside in southern border states (border with Mexico) compared to individuals not from southern border states. A random sample of individuals is taken from southern border states as well as non-border states; the total sample size is 1000. Respondents are asked to rate immigrants on a 100-point feeling thermometer, where 100 denotes most favorable attitudes; 50 denotes "middle of the road" attitudes; and 0 denotes least favorable attitudes. The following results are obtained:

For respondents from southern border states: $\overline{X} = 46$
For respondents from non-border states: $\overline{X} = 50$
The $SEDM = 2.40$

a. What is the null hypothesis? (3 points)
b. What is the alternative hypothesis and why is it stated this way? (6 points)
c. Using $\alpha = .05$, can you reject the null hypothesis? Why/how did you come to this decision? (8 points)
d. If $\alpha = .01$, can you reject the null? Why or why not? (5 points)

4. A researcher is interested in the relationship between the inflation rate in a country and strike volume in a country. Strike volume is measured as days lost due to industrial disputes per 1000 wage salary earners. She collects data over a number of years and records the strike volume and the inflation rate. A scatter plot of the data reveals the relationship shown in the graph below.

a. Please describe the main features of the graph. What is the dependent variable and what is the independent variable in this research question? (6 points)



year	strikes	inflation
1951	773	7.9
1952	752	9.2
1953	117	5
1954	95	0
1955	336	2.6
1956	70	4.3
1957	135	4.1
1958	187	4.5
1959	181	0
1960	115	.4
1961	589	2.7
1962	160	4.2
1963	354	2.5
1964	794	6.7
1965	796	5
1966	1129	3.1
1967	262	3.2
1968	574	4.7
1969	1304	7.4
1970	1405	8.2
1971	379	9
1972	284	8.7
1973	277	11.4
1974	720	17
1975	390	20.9
1976	1032	18
1977	571	13.6
1978	765	7.6
1979	1752	13.2
1980	480	18.2
1981	503	20.4
1982	504	17.1
1983	382	10.5
1984	472	8.6
1985	521	5.4

Formulae that may be of interest:

Mean:

$$\overline{X} = \frac{\sum_i^n (X_i)}{n} \tag{1}$$

Variance

$$s^2 = \frac{\sum_i^N (X_i - \overline{X})^2}{n-1} \tag{2}$$

Standard Deviation

$$s = \sqrt{(s^2)} \tag{3}$$

z-score

$$z = \frac{X_i - \overline{X}}{s} \tag{4}$$

Median

$$M = \frac{N+1}{2} \tag{5}$$

One-sample *t*

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} \tag{6}$$

Two-group *t*

$$t = \frac{\overline{X}_{Dem} - \overline{X}_{Rep}}{SEDM} \tag{7}$$

t-ratio for regression

$$t = b/s.e.(b) \tag{8}$$