Stats 102
Prof. Green

Review Problems for the Midterm

Disclaimer: The following problems may help you prepare for the midterm exam. This study guide may not provide coverage of every topic that will appear on the exam. When studying, make use of the website's detailed explanation of statistical topics.

1. Provide brief definitions of the following terms: dichotomous data, experiment, sampling distribution, central limit theorem, median vs. mean

2. A sample of 100 nation-states is drawn, and the number of chemical weapons is counted for each state. States fall into the categories in the table below.

Draw a density histogram to represent these data.

Weapons	**0-99**	**100-199**	**200-299**	**300-399**	**400-599**
Percentage of states	**10**	**25**	**35**	**15**	**15**

3. A national survey reports the following results based on a simple random sample of the US adult population: 40% of the 800 people interviewed believe that President Clinton should be impeached. A week earlier, a survey using the same methodology found that 35% of the 800 people interviewed had expressed this view. Test the hypothesis that true opinion had in fact remained unchanged during this period. Can you reject the null hypothesis at the 5% significance level?

4. Suppose that 40% of all American adults would be willing to endorse public financing of political campaigns. That is, the population proportion p=.4. Suppose a simple random sample of adults is conducted, such that the sampling distribution is approximately normal with a mean of .4 and a standard deviation of 0.183. Using this information, find the probabilities of the following events.

a). At least half of the sample answers endorses public funding.

b). Less than 35% of the sample endorses public funding.

c). The sample proportion is between .37 and . 46.

5. In a major city, the mean housing price was measured for each of the preceding 8 years. At the same time, the net change in population (inflow-outflow) of the city was measured.

Year	Housing price	net population change (in 1000's)
1972	52,000	33
1973	53,000	34
1974	55,000	21
1975	56,000	20
1976	52,000	38
1977	53,000	42
1978	54,000	50
1979	56,000	51

a). Calculate the correlation between the two variables.

b). How would the correlation change if housing prices were measured in Canadian dollars, where $0.67US = $1 CND?

6. Policymakers worry that Americans don't save enough of their income. An experiment is performed to assess how offering families tax incentives affects their rate of saving. Five clusters of families are randomly assigned to different tax incentive schemes. The independent variable is the average tax break per $1000 of income under each level of the experimental plan. The amount families in a given group choose to save per $1000 of income is the dependent variable.

tax incentives	savings rate
0	8
5	8
10	11
15	12
20	12

Regress savings on tax incentives. Interpret the coefficients and the R-square. What are the implications of these results?

7. In the aftermath of an election, pundits interpret the election of a liberal candidate (Smith) as an endorsement of her pro-choice stance on abortion. Here is a three way table of how Democrats and Republicans voted, broken down by their stance on abortion. Entries in each cell are the number of people in each category surveyed during an exit poll.

	Democrats		Republicans	
	Pro-choice	Pro-life	Pro-choice	Pro-life
Voted for Smith	500	150	50	200
Voted against Smith	100	30	150	600

Percentage the table intelligently and interpret the results. Construct a crosstabulation of vote by stance on abortion. Is there a relationship between the two variables? How does this relationship differ in its implications from the one depicted in the three-way table above?

8. Explain how the conditional probability P(A|B) differs from the unconditional probabilities P(A) and P(B). Contrast P(A|B) with P(B|A).

9. Suppose that you are on an administrative panel charged with deciding whether a student cheated on a multiple choice exam. Suppose that the person charged with cheating has had no previous record of cheating, so that the prior probability of cheating P(C)=.01. The reason to suspect cheating in this case is that the alleged cheater had a large number of right and wrong answers that were identical to those of another test-taker. If cheating were not occurring, the probability that this would occur is .01. On the other hand, if cheating were going on, duplicate answers would occur 60% of the time. Given this information, what probability do you assign the hypothesis that the accused in fact cheated?

10. You are a pollster seeking to measure the percentage of the teenagers who are smoke. You wish to conduct a poll with a standard deviation (standard error) of 5 percentage-points. Explain the reasoning that goes into your decision of how many people need to be sampled in your poll.

11. Suppose state governors, on average, appoint 2 Supreme Court justices per 4 years, with a standard deviation of 1.2 justices. How many justices do governors, on average, appoint during a single year? What is the standard deviation of the yearly rate?