Statistics 7 NAME: KEY

Midterm Exam 1 Last six digits of Student ID#:

Make sure you have 5 pages. You may use one page of notes (both sides) and a calculator. For multiple choice questions, *circle* the best answer. There are 20 questions, each worth 5 points. If you need extra space use the back of the page, but make sure to point to it.

Questions 1 and 2: A Roper poll conducted in the first week of October, 2009 asked a random sample of 1003 adults in the United States "Would you say that now is a good or a bad time to invest in the stock market?" 42% of the respondents said "good time."

1. What is the (conservative) margin of error for this poll? Show your work.

$$\frac{1}{\sqrt{1003}} \times 100\% = 3.2\% \text{ (or you can leave it as .032)}$$

You can also round off to .03 or 3%.

2. Provide a 95% confidence interval for the percent of all adults in the United States who thought that it was a good time to invest in the stock market. Show your work.

$42\% \pm 3.2\%$
38.8% to 45.2%
You can round off: 39% to 45%

3. An NBC News/ Wall Street Journal poll conducted in April, 2009 asked a random sample of 500 adults in the United States "Would you approve or disapprove of a proposal that would require companies to reduce greenhouse gases that cause global warming, even if it would mean higher utility bills for consumers to pay for the changes?" 53% of the respondents said "approve." The poll reported a margin of error of 4.4%. Which of the following statements (applied to April 2009) is correct based on the results of this poll?
   A. We can be certain that the percent of all adults in the US who would approve of such a proposal falls in the interval 48.6% to 57.4%.
   B. We can be 95% confident that the percent of adults in the US who would approve of such a proposal was between 48.6% and 57.4%.
   C. We can be 4.4% confident that the percent of adults in the US who would approve of such a proposal was 53%.
   D. The chance is 95% that at least 53% of adults in the US would approve of such a proposal.

4. A deterministic relationship between two quantitative variables X and Y is one in which:
   A. If we know the value of X we can completely determine the value of Y.
   B. We know that X is a determining factor in causing Y.
   C. We know that Y is a determining factor in causing X.
   D. The magnitude of the correlation between X and Y cannot be determined.

Questions 5 to 8: One of the teams in the first discussion asked students to specify their gender (male or female), and to answer the question "Have you ever been pulled over by an officer?" The results of a chi-square test for this situation are shown below. For example, there were 7 males who responded "no" and the expected count for that cell is 9.04. That cell contributes 0.459 to the chi-square statistic, which has a value of 1.282 and a p-value of 0.258.

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts
           Ever Pulled Over?
              No    Yes  Total
    Male       7     15     22
            9.04  12.96
           0.459  0.320
  Female      16     18     34
           13.96  20.04
           0.297  0.207
   Total      23     33     56

Chi-Sq = 1.282, DF = 1, P-Value = 0.258
```

5. Write the null and alternative hypotheses for this situation. You can assume that these students are representative of all UCI students.

   Null hypothesis: There is no relationship between gender and whether or not one has been pulled over by an officer, for the population of UCI students.

   Alternative hypothesis: There is a relationship between gender and whether or not one has been pulled over by an officer, for the population of UCI students.

6. Show how the expected count of 12.96 for "Male, Yes" was computed.

$$\frac{(row\ total)(column\ total)}{total} = \frac{(22)(33)}{56} = 12.96$$

7. Which of the following is a valid conclusion for this chi-square test?
   A. There is no relationship between gender and being pulled over for the population of UCI students.
   B. There is a relationship between gender and being pulled over for the population of UCI students.
   C. The relationship between gender and being pulled over is not statistically significant.
   D. The relationship between gender and being pulled over is statistically significant.

8. In this situation, the *expected counts* are calculated by assuming:
   A. there is no relationship between gender and being pulled over, for the population of UCI students.
   B. there really is a relationship between gender and being pulled over, for the population of UCI students.
   C. the alternative hypothesis is true.
   D. UCI students are not pulled over any more often than other people are pulled over.

9. Which of the following is the best example of a population from which a cluster sample would be easier to obtain than a simple random sample?
   A. Registered drivers in a certain state.
   B. Employees of a large company.
   C. Members of a national organization.
   D. Passengers who will be flying in the upcoming week on a certain airline.

Questions 10 to 13: A group of adults aged 20 to 80 were tested to see how far away they could first hear an ambulance coming towards them. A regression equation describing the relationship between y = distance (in feet) and x = age was found to be:

$$\text{Distance} = 600 - 3 \times \text{Age}$$

10. In this situation, what is the explanatory variable and what is the response variable?

    Explanatory variable is ____Age_____

    Response variable is ____Distance at which one can first hear an ambulance coming__

    (It is acceptable but not ideal to just say "distance" for the response variable.)

11. Predict the distance at which someone who is 30 years old would first hear the ambulance. Show your work.

$$\text{Distance} = 600 - 3(30) = 600 - 90 = 510 \text{ feet}$$

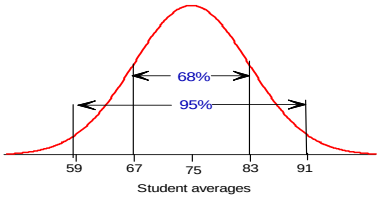
12. Explain what it would mean in this situation if a 30-year-old had a residual of 10.

    Residual = observed – predicted = 10, which means that the observed value is 10 feet higher than the predicted value. Therefore, this person could hear the ambulance at 520 feet.

13. One interpretation of the slope in this situation is
    A. As someone's age goes up by 1 year, distance at which they can hear the ambulance is predicted to go up by 3 feet.
    B. As someone's age goes up by 1 year, distance at which they can hear the ambulance is predicted to go down by 3 feet.
    C. As someone's age goes up by 1 year, distance at which they can hear the ambulance is predicted to go up by 600 feet.
    D. As someone's age goes up by 1 year, distance at which they can hear the ambulance is predicted to go down by 600 feet.

Questions 14 and 15: A Statistics professor has taught thousands of students, and has found that the student averages at the end of the quarter (based on all homework and exam scores) have approximately a bell-shaped distribution with mean of 75 and standard deviation of 8.

14. Draw a picture of the distribution of student averages for this professor and indicate the locations (including numerical values) of the mean, the interval that covers the middle 68% of the values and the interval that covers the middle 95% of the values.



The Empirical Rule tells us that 68% fall within 1 standard deviation of the mean, which is 75 ± 8, or 67 to 83, and 95% fall within 2 standard deviations of the mean, which is 75±16, or 59 to 91.

15. What is the standardized score for a student who has an average of 91? About what percent of students have *higher* averages than this student? (*Hint:* Use your picture above.)

Standardized score = $\frac{91-75}{8} = \frac{16}{8} = 2$ % who are higher = ___2.5%___

16. Below is a stem-and-leaf plot for the times (in minutes) it took 15 students to finish an exam in a 50-minute class. Create a five-number summary for the data. (Put it to the right of the stem-and-leaf plot.)

|2|
|2| 89
|3| 12344
|3|
|4| 113
|4| 679
|5| 00
|5|

41
32 ***47***
28 ***50***

17. The stem-and-leaf plot in Question #16 repeated each of the numbers 2 to 5 twice as stem values. In general, which of the following is *not* an allowable number of times to repeat numbers as stem values?

A. 3
B. 5
C. 10
D. All of the above are allowable numbers of times to repeat numbers as stem values.

18. Eighty individuals who wished to lose weight were randomly divided into two groups of 40. One group was given an exercise program to follow while the other group was given a special diet. After three months, the researcher compared mean weight losses in the two groups. What type of study is this?
    A. Observational study
    B. Repeated measures study
    C. Matched pairs study
    D. Randomized experiment

19. The table below was presented in class, showing what conclusions can be made with randomized experiments and observational studies, when the sample used for the study represents the population for the question of interest, or does not represent that population. One of the following sets of conclusions goes into each of the four cells of the table. Indicate which conclusion goes into which cell by circling the appropriate letter in that cell. The possible sets of conclusions are:

    A. Can conclude causal relationship, but cannot extend results to the population
    B. Cannot conclude causal relationship, and cannot extend results to the population
    C. Can conclude causal relationship, and can extend results to the population
    D. Cannot conclude causal relationship, but can extend results to the population

	Sample represents population for question of interest	Sample doesn't represent population for question of interest
Randomized Experiment	***C***	***A***
Observational Study	***D***	***B***

20. If a relationship seen in sample results is "statistically significant" it means that
    A. there is an important relationship in the sample results.
    B. there is an important relationship in the population.
    C. the relationship observed in the sample was unlikely to have occurred if there is no relationship in the population.
    D. there is an important relationship in the population that wasn't reflected in the sample results.