

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: 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.”

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

$$\frac{1}{\sqrt{500}} \times 100\% = 4.5\% \text{ (or you can leave it as .045)}$$

2. Provide a 95% confidence interval for the percent of all adults in the United States who would have approved of such a proposal at that time, if asked. Show your work.

$$53\% \pm 4.5\% \\ 48.5\% \text{ to } 57.5\%$$

3. A Gallup poll conducted in May, 2009 asked a random sample of 1015 adults in the United States “Right now, do you think the state of moral values in the country is getting better or getting worse?” and 71% responded “getting worse.” Which of the following statements (applied to May 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 thought moral values were getting worse was in the interval 68% to 74%.
- B. We can be 95% confident that the percent of adults in the US who thought moral values were getting worse was between 68% and 74%.**
- C. We can be 3% confident that the percent of adults in the US who thought moral values were getting worse was 71%.
- D. The chance is 95% that at least 71% of adults in the US thought moral values were getting worse.
4. A deterministic relationship between two quantitative variables X and Y is one in which:
- A. We know that X is a determining factor in causing Y.
- B. We know that Y is a determining factor in causing X.
- C. The magnitude of the correlation between X and Y cannot be determined.
- D. If we know the value of X we can completely determine the value of Y.**

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 in a car accident?” 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 9.28. That cell contributes 0.560 to the chi-square statistic, which has a value of 1.575 and a p-value of 0.209.

Expected counts are printed below observed counts			
Chi-Square contributions are printed below expected counts			
Ever in an Accident?			
	No	Yes	Total
Male	7	16	23
	9.28	13.72	
	0.560	0.379	
Female	16	18	34
	13.72	20.28	
	0.379	0.256	
Total	23	34	57

Chi-Sq = 1.575, DF = 1, P-Value = 0.209

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 ever been in an accident, for the population of UCI students.

Alternative hypothesis: There is a relationship between gender and whether or not one has ever been in an accident, for the population of UCI students.

6. Show how the expected count of 13.72 for “Male, Yes” was computed.

$$\frac{(\text{row total})(\text{column total})}{\text{total}} = \frac{(23)(34)}{57} = 13.72$$

7. Which of the following is a valid conclusion for this chi-square test?
- There is no relationship between gender and being in an accident for the population of UCI students.
 - There is a relationship between gender and being in an accident for the population of UCI students.
 - The relationship between gender and being in an accident is not statistically significant.**
 - The relationship between gender and being in an accident is statistically significant.
8. In this situation, the *expected counts* are calculated by assuming:
- there is no relationship between gender and being in an accident, for the population of UCI students.**
 - there really is a relationship between gender and being in an accident, for the population of UCI students.
 - the alternative hypothesis is true.
 - UCI students are not in accidents any more often than other people are in accidents.

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?
- Passengers who will be flying in the upcoming week on a certain airline.*
 - Registered drivers in a certain state.
 - Employees of a large company.
 - Members of a national organization.

Questions 10 to 13: The following regression equation was computed based on 20 cities in the United States for which y = average January temperature and x = latitude were known:

$$\text{Average January temperature} = 126 - 2.34 \times \text{latitude}$$

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

Explanatory variable is latitude

Response variable is Average January temperature

11. Predict the average January temperature for Houston, which has latitude of 30. Show your work.

$$\text{Avg. Jan. temp} = 126 - 2.34(30) = 55.8 \text{ degrees}$$

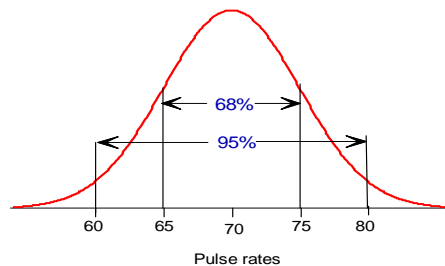
12. In this situation Houston actually has a residual of -5.5 degrees. Explain what this means.

Residual = observed – predicted = -5.5 degrees, which means that the observed value for Houston is 5.5 degrees lower than the predicted value. Therefore, the actual average January temperature for Houston is $55.8 - 5.5 = 50.3$ degrees.

13. One interpretation of the slope in this situation is
- If one city has latitude of 10 degrees more than another city, it is predicted to have an average January temperature 23.4 degrees lower than the first city.***
 - If one city has latitude of 10 degrees more than another city, it is predicted to have an average January temperature 23.4 degrees higher than the first city.
 - If one city has latitude of 10 degrees more than another city, it is predicted to have an average January temperature that differs by 23.4 degrees from that of the first city, but we don't know if it will be higher or lower.
 - A city with latitude 0 is predicted to have an average January temperature of 126 degrees.

Questions 14 and 15: Based on measurements from several thousand college students, it is known that pulse rates for students have approximately a bell-shaped distribution with mean of 70 and standard deviation of 5.

14. Draw a picture of the distribution of student pulse rates 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 70 ± 5 , or 65 to 75, and 95% fall within 2 standard deviations of the mean, which is 70 ± 10 , or 60 to 80.

15. What is the standardized score for a student who has a pulse rate of 80? About what percent of students have *higher* pulse rates than this student? (*Hint:* Use your picture above.)

$$\text{Standardized score} = \frac{91 - 75}{8} = \frac{16}{8} = 2 \quad \% \text{ who are higher} = \underline{\quad 2.5\% \quad}$$

16. Below is a stem-and-leaf plot for the times (in minutes) it took 15 students to finish an exam in an 80-minute class. Create a five-number summary for the data. (Put it to the right of the stem-and-leaf plot.)

5		
5 89		
6 12344	70	
6	62	77
7 013	58	80
7 679		
8 00		
8		

17. The stem-and-leaf plot in Question #16 repeated each of the numbers 5 to 8 twice as stem values. In general, which of the following is *not* an allowable number of times to repeat numbers as stem values?
- A. 10
 B. 5
 C. 3
 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?
- Observational study
 - Randomized experiment**
 - Repeated measures study
 - Matched pairs study
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:
- Cannot conclude causal relationship, but can extend results to the population
 - Can conclude causal relationship, but cannot extend results to the population
 - Cannot conclude causal relationship, and cannot extend results to the population
 - Can conclude causal relationship, and can extend results to the population

	Sample represents population for question of interest	Sample doesn't represent population for question of interest
Randomized Experiment	<i>D</i>	<i>B</i>
Observational Study	<i>A</i>	<i>C</i>

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