22S:30/105, Statistical Methods and Computing
Spring 2012, Instructor: Cowles
Midterm 3

Show your work on any problems that involve calculations.

Name: _____ Course no. (30 or 105) _____

1. An actuarial science instructor at an Iowa university wished to determine the proportion of Iowa high school seniors who are aware of the career of actuary. He identified a simple random sample of 25 Iowa high school seniors and asked each one whether he or she had heard of actuaries and knew that there was such a career. Eleven of the high school seniors answered "yes" and 14 answered "no."

   (a) What is the population of interest to the instructor? (circle one)
      i. the proportion of all Iowa high school seniors who are aware of the career of actuary
      ii. the proportion of the 25 Iowa high school seniors who answer "yes"
      iii. all Iowa high school seniors
      iv. the 25 Iowa high school seniors whom the instructor surveyed
      v. none of the above

   (b) What is the parameter of interest to the instructor? (circle one)
      i. the proportion of all Iowa high school seniors who are aware of the career of actuary
      ii. the proportion of the 25 Iowa high school seniors who answered "yes"
      iii. all Iowa high school seniors
      iv. the 25 Iowa high school seniors whom the instructor surveyed
      v. none of the above

   (c) The instructor wished to calculate a point estimate and 90% confidence interval for the quantity of interest. Check the rules of thumb to determine whether the normal approximation will give a trustworthy confidence interval.

(d) Calculate the point estimate and 90% interval two ways. In both cases, show your work and give the numeric results.

i. using the normal approximation

ii. using the plus-four method

(e) The symbol used in this class for the quantity that you are 90% confident lies in your interval is: (circle one)

i. $\bar{x}$

ii. μ

iii. $\hat{p}$

iv. p

v. none of the above

2. Following each of the following descriptions of a research study, circle the statistical procedure that is most likely to be appropriate.

   (a) The actuarial science instructor in problem 1 wishes to test whether the proportion of Iowa high school seniors who are aware of the career of actuary is less than 0.5.

      i. one-sample z test
      ii. paired sample z test
      iii. two-independent-sample z test
      iv. one-sample t test
      v. paired sample t test
      vi. two-independent-sample t test
      vii. chi-square test

   (b) A veterinarian wishes to test whether the population mean of systolic blood pressure is different in dogs from in cats. He randomly selects 15 dogs and 15 cats from among his patients and measures the systolic blood pressure of each animal.

      i. one-sample z test
      ii. paired sample z test
      iii. two-independent-sample z test
      iv. one-sample t test
      v. paired sample t test
      vi. two-independent-sample t test
      vii. chi-square test

   (c) A high school French teacher wishes to determine whether male or female high school students are better in French courses. She random selects one male student and one female student from each French class taught at her school, and obtains the scores for these students on the final exam in their French classes.

      i. one-sample z test
      ii. paired sample z test
      iii. two-independent-sample z test
      iv. one-sample t test
      v. paired sample t test
      vi. two-independent-sample t test
      vii. chi-square test

3. Biologists wished to study whether the size of ants affects how far they move away from their home ant hill as they forage for food. They trapped ants at two different distances from their ant hills – 5 meters and 10 meters – and then measured the diameter of each ant's head. (Head diameter is a good indicator of the overall size of the ant.) They classified the ants into 4 categories based on size. Below is a frequency table of the counts of each size of ant at 5 meters and at 10 meters from their hills.

class	distance		
Frequency Percent Row Pct Col Pct	5	10	Total
<37	51 14.33 71.83 23.50	20 5.62 28.17 14.39	71 19.94
37-38	61 17.13 58.10 28.11	44 12.36 41.90 31.65	105 29.49
39-40	62 17.42 57.94 28.57	45 12.64 42.06 32.37	107 30.06
>40	43 12.08 58.90 19.82	30 8.43 41.10 21.58	73 20.51
Total	217 60.96	139 39.04	356 100.00

Statistics for Table of class by distance

Statistic	DF	Value	Prob
Chi-Square	3	4.4262	0.2190
Likelihood Ratio Chi-Square	3	4.5770	0.2055
Mantel-Haenszel Chi-Square	1	0.5786	0.4469
Phi Coefficient		0.1115	
Contingency Coefficient		0.1108	
Cramer's V		0.1115	

Sample Size = 356

(a) The biologists wish to use their data to test the null hypothesis that the proportion of ants found 10 meters away from the anthill is the same in all 4 sizes classes. This null hypothesis could be written as follows, where p_1 represents the proportion of ants in the smallest class found 10 meters away, p_2 represents the proportion in the second smallest class found 10 meters away, etc.

$$H_0 : p_1 = p_2 = p_3 = p_4$$

Write the corresponding alternative hypothesis.

(b) At the .05 signficance level, does the data provide strong enough evidence to reject the null hypothesis? Refer to specific SAS output to justify your answer.

(c) Should the biologists proceed to try to determine which size classes are significantly different from each other with respect to the proportion of ants that travel 10 meters away from the hill? Why or why not?

.

(d) Compute the *expected* number of ants in the `<37` class to be found 10 meters away from the hill if the null hypothesis is true. Show your work.

4. Researchers wish to perform a hypothesis test to test the alternative hypothesis that the mean score on a particular intelligence test is greater then 100 in the population of all Iowa City adult residents, versus the null hypothesis that the population mean is less than or equal to 100.

$$H_0 : \mu \leq 100$$

$$H_A : \mu > 100$$

Suppose that individual scores on this intelligence test are known to follow an approximately normal distribution with standard deviation $\sigma = 10$ points.

(a) Suppose that the researchers want to do their hypothesis test at significance level $\alpha = 0.01$, and that they have give the test to a simple random sample of 50 adult Iowa Citians. How large a value of the sample mean $\bar{x}$ will be required in order to reject the null hypothesis? Numeric answer; show your work.

(b) With the above significance level and sample size, how much power will the test have against the alternative that $\mu = 110$? Numeric answer; show your work.

(c) Would the power in the previous question be larger or smaller if the researchers had a simple random sample of size 100 rather than 50?