Econometric Analysis
Homework 3

Given: Oct 26, 2007
Due: October 15, 2007

Jin-Lung Lin

1. The Environmental Protection Agency and the University of Florida recently cooperated in a large study of the possible effects of trace elements in drinking water on kidney stone disease. The accompanying table presents data on age, amount of calcium in home drinking water(measured in parts million), and smoking activity. These data were obtained from individuals with recurrent kidney stone problems, all of whom lived in the Carolinas and the Rocky Mountain states.

	Carolinas	Rockies
Sample Size	467	191
Mean age	45.1	46.4
Standard deviation of age	10.2	9.8
Mean calcium component (ppm)	11.3	40.1
Standard deviation of calcium	16.6	28.4
Proportion now smoking	.78	.61

   a. Estimate the average calcium concentration in drinking water for kidney stone patients in the Carolinas. Place a bound on the error of estimation.

   b. Estimate the difference in mean ages for kidney-stone patients in the Carolinas and in the Rockies. Place a bound on the error of estimation.

   c. Estimate and place a two-standard-deviation bound on the difference in proportions of kidney-stone patients from the Carolinas and Rockies who were smokers at the time of study.

2. In a study of the relationship between birth order and college success, an investigator found that 126 in a sample of 180 college graduates were firstborn or only children; in a sample of 100 nongraduates of comparable age and socioeconomic background, the number of firstborn or only children was 54. Estimate the difference in the proportions of first-born or only children for two populations from which these samples were drawn. Give a bound for the error of estimation.

3. Suppose that the random variable Y is an observation from normal distribution with unknown mean μ and variance 1.

a. Find a 95 % confidence interval for μ.

b. Find a 95 % upper confidence limit for μ.

c. Find a 95 % lower confidence limit for μ.

4. Athletes at major colleges graduated, on the whole, at virtually the same rate as other students," according the National Collegiate Athletic Association (NCAA). The survey involved students who had entered college before the NCAA's Proposition 48 took effect in 1985. (Proportion 48 requires that athletes maintain at least a 2.0 grade point average on a scale of 4.0.) Suppose that, in a new poll of 500 athletes at major colleges, the number graduating was 268.

   a. Find a 98 % confidence interval for p, the proportion of athletes at major colleges who graduate.

   b. Does the interval include the value p=.51, the graduation rate prior to proportion 48? What do you conclude?

5. Alice Chang and her colleagues tested 559 high school students using leisure-interest checklist (LIC) survey. The means and standard deviation for each of the seven LIC factors are given in the following table for n=252 male students and n=307 female students.

	Males		Females	
LIC Factor Scale Activity	Mean	s	Mean	s
Hobbies	10.06	6.47	13.64	7.46
Social fun	22.05	5.12	25.96	5.07
Sports	13.65	4.82	9.88	4.41
Cultural	11.48	5.69	13.21	5.31
Trips	6.90	3.41	6.49	2.97
Games	4.95	3.29	3.85	2.49
Church	2.15	2.10	3.00	2.26

   a. Give a 95% confidence interval for the difference in the means for male and female students for the cultural activity scale. Interpret this interval.

   b. Find a 90% confidence for the difference in the means for male and female students for the social fun factors scale activity.

   c. Do the means for males and females for the two LIC activities considered in (a) and (b) appear to differ?

6. Scholastic Aptitude Test (SAT) scores, which have fallen slowly since the inception of the test, have now begun to rise. Originally, a score of 500 was intended to be average. The mean scores for 1991 were approximately 422 for the verbal test and 474 for the mathematics test. A random sample of test scores of 20 seniors from a large urban high school produced the means and standard deviations listed in the accompanying table.

	Verbal	Mathematics
Sample mean	419	455
Sample standard deviation	57	69

   a. Find a 90 % confidence interval for the mean verbal SAT scores for urban high school seniors.

   b. Does the interval that you found in (a) include the value 422, the true mean verbal SAT score for 1991? What can you conclude?

   c. Construct a 90 % confidence interval for the mean mathematics SAT score for the urban high school seniors. Does the interval include 474, the true mean mathematics score for 1991? What can you conclude?

7. Do SAT scores for high school students differ depending on the students' intended field of study? Fifteen students intended to major in engineering were compared with 15 students who intended to major in language and literature. Given in the accompanying table are the means and standard deviations of the scores on the verbal and mathematics portion of the SAT for the two groups of students.

	Verbal		Math	
Engineering	$\bar{y}$=446	s=42	$\bar{y}$=548	s=57
Engineering	$\bar{y}$=534	s=45	$\bar{y}$=517	s=52

   a. Construct a 95 % confidence interval for the difference in average verbal scores of students majoring in engineering and of those majoring in language/literature.

   b. Construct a 95 % confidence interval for the difference in average math scores of students majoring in engineering and of those majoring in language/lierature.

   c. Interpret the results obtained in (a) and (b).

   d. What assumptions are necessary for the methods used previously to be valid?

8. The sample mean $\bar{Y}$ is a good point estimator of the population mean μ. It can also be used to predict a future value of Y independently selected from the population. Assume that you have a sample mean $\bar{Y}$ and variance S^2

based on a random sample n measurements from a normal distribution. Use Student's t to form a pivotal quantity to find a prediction interval for some new value of Y, say, Y_p, to be observed in the future.