Lab 4: Sampling & Confidence Intervals   **NAME**:_____

Main Objectives:

- Help you understand 1) sampling theory and 2) confidence intervals
- Demonstrate the importance of avoiding bias so that a sample is obtained that accurately estimates species occurrences.

PART 1: SAMPLING

Why Sample?

Sampling methods are invaluable for numerous biological investigations. Such methods are used to determine the efficacy of new medicines, the responses of cells to various treatments, and the structure of a natural community. It would be extremely time consuming, for example, to count and measure every individual of each species within a community in order to determine the abundances and distributions of each species within the community. Sampling methods enable us to estimate reliable information by use of samples. However, it is critical that the samples be taken without bias and in sufficiently large number so that the resulting data can be summarized to give valid estimates of the desired parameters.

1. Using the Urban Land Survey data you will examine some sample statistics and compare them to the population parameters. To do this first calculate the population mean and standard deviation for the *value* column (the 86 cases will be our population). Then take <u>5</u> random samples of <u>30</u> elements from this column. To do this in Excel, first makes sure that the <u>Analysis ToolPAK</u> is activated. Then go to the <u>DATA menu</u>, choose DATA ANALYSIS and select *sampling*. When the sampling box appears fill in the input range by highlighting the entire column for "value" in the data set. Select *Random* sampling (put 30 in the box that says *number of samples* – which really means just "*n*"), and select a cell to begin your output range. Do this <u>5</u> times (best to put the sample columns next to each other). Calculate the mean and standard deviation for each sample and then calculate the mean and standard deviation (Tools → Data Analysis → Descriptive statistics) of the sample means and sample standard deviations. Fill in the following table. **Discuss the variation you see in the sample statistics and compare the mean of the sample means and the mean of the sample standard deviations to the population parameters (mean and standard deviation).** (1 point)

	Mean	Standard Deviation
Entire population		
Sample 1		
Sample 2		
Sample 3		
Sample 4		
Sample 5		
Summary of the sample statistics		

2. As in question 1, use the sampling function on Excel to sample the "value" variable. This time instead of collecting 5 samples of 30 elements, you will collect increasing large samples to monitor the change in the sample statistics with increasing sample size. Fill in the following table. **Discuss the variation you see in the sample statistics and compare the mean of the sample means and the mean of the sample standard deviations to the population parameters (mean and standard deviation). (1 point)**

	Mean	Standard Deviation
Of the population		
Sample 1 (5 elements)		
Sample 2 (10 elements)		
Sample 3 (30 elements)		
Sample 4 (50 elements)		
Sample 5 (70 elements)		
Summary of the sample statistics		

3. Collecting a *stratified random sample*. You are going to examine property value, stratifying your sample by the age class of the property. The properties are grouped into age classes from 1 to 4. Collect a sample of properties (6 from each age group). To randomly select the samples use the random number table on page 304 of your text (Appendix B). Point to a spot on the table and make this your starting point. Since these are five digit numbers (and our cases only range from 1 to 86), use the first two or last two digits of the five digit number to relate to your case numbers. Then moving either up or down in the table select the *case* from the Urban Land Survey according to the random number (if the number is over 86 skip to the next and if you repeat a number skip to the next random dumber). If you have already chosen 6 cases from one of the age classes, then reject that random number and move to the next one. When you are finished you will have chosen 6 cases (and recorded the property values) from each property age class. Show your sample of <u>values</u> below in a <u>*well-labeled table*</u>. How does the mean and standard deviation of this sample compare to the population parameters calculated in question 1? (2 point)

PART 2: CONFIDENCE INTERVALS

4. What Z-values correspond to the following significance levels for confidence interval tests: (2 points)

    a) 95% confidence limits

    b) 99% confidence limits

    c) 99.5% confidence limits

5. Construction of confidence intervals for samples (using sample variance): (2 points)

For the Urban Land Survey Data construct the following confidence limits

    a) 95% confidence interval around the sample mean of **area**.

    b) 95% confidence interval around the sample mean of **value**.

    c) 99% confidence interval around the sample mean of **value**.

    d) What happened when you changed the confidence level from 95% to 99% for the estimate of the mean of value?

    e) Construct the 90% confidence limits for the following samples (n = 18) and comment on the nature of the sampled distributions (i.e., is there a lot of variation around the mean).

Sample 1:	mean = 100	Sample 2:	mean = 100
	std. dev. = 3.6		std. dev. = 12.4
	n = 18		n = 18

6. The following measurements of lead particulates have been measured in the air at three locations in northern Illinois. The EPA has set a safe maximum level at 5.1. For each location determine the mean value, standard deviation, upper and lower 95% confidence interval. Do the lead concentrations at the sites exceed the EPA's standards? Explain your answer. [SPSS may be used for this problem]. **(2 points)**

	Site 1	Site 2	Site 3
	6.86	5.55	6.61
	4.05	3.03	1.01
	5.92	4.71	4.74
	9.00	7.48	10.91
	6.91	5.60	6.73
	12.29	10.45	17.49
	10.87	9.16	14.63
	5.14	4.01	3.18
	9.49	7.92	11.89
	5.48	4.31	3.85
	7.50	6.13	7.90
	0.22	0.01	0.01
	5.61	4.43	4.12
	7.66	6.28	8.23
	10.97	9.25	14.83
	6.09	4.86	5.08
	9.39	7.83	11.67
	6.49	5.22	5.88
	4.20	3.16	1.31
	7.59	6.21	8.08
	4.57	3.51	2.06
	4.80	3.71	2.49
	7.52	6.15	7.94
	1.98	1.16	0.01
	7.72	6.33	8.33
	4.71	3.62	2.32
	5.18	4.04	3.26
	8.02	6.61	8.94
	1.63	0.85	0.01
	8.70	7.20	10.31
mean			
std. dev			
lower 95% conf. limit			
upper 95% conf. limit			