Homework 4

Due: Friday, Nov. 18, 2009

1. The spread sheet "birth weight.xlsx" contains the birth weights from 1000 consecutive deliveries at Boston City Hospital. Consider these 1000 birth weights as our population in consideration.

   (a) Generate the sampling distribution of $\bar{X}$ over 200 samples of size 10 and plot the distribution you generate. (Hint: You can use whatever programming language you feel comfortable with. Use a uniform random generator to obtain 200 random samples with each random sample having 10 birth weights. For each random sample, compute its sample mean. You will have a total of 200 sample means and plot the histogram of these 200 numbers.)

   (b) For the 200 random samples you obtained in (a), generate the sampling distribution of the sample median.

   (c) For the 200 random samples you obtained in (a), generate the sampling distribution of the average of the smallest and largest observation.

   (d) What conclusion can you draw by comparing the three sampling distributions you generate in (a), (b), and (c)?

2. Consider the 1000 birth weights as our population again. Generate the sampling distribution of the sample mean using 200 samples of different sizes.

   (a) $n = 10$

   (b) $n = 20$

   (c) $n = 30$

   (d) What conclusion can you draw by comparing the three sampling distributions you generate in (a), (b), and (c)?

3. (Illustration of the central limit theorem) Consider the 1000 birth weights as our population again. Generate the sampling distribution of the sample mean over 200 samples of different sizes.

   (a) $n = 1$

   (b) $n = 5$

   (c) $n = 10$

   (d) Can you see that your sampling distributions of the sample mean can be more closely approximated by a Gaussian distribution when n increases? Explain why.

4. The 1999-2000 National Health and Nutrition Examination Survey was used to estimate dietary intake of 10 key nutrients. One of those nutrients was calcium (mg). They found in all adults 60 years or older a mean daily calcium intake of 721 mg with a standard deviation of 454. Using these values for the mean and standard deviation for the U.S. population, find the probability that a random sample of size 50 will have a mean:

   (a) Greater than 800 mg

(b) Less than 700 mg

(c) Between 700 and 850 mg

5. Suppose a population consists of the following values: 1,3,5,7,9. Construct the sampling distribution of $\bar{x}$ based on samples of size 2 selected without replacement. Find the mean and variance of the sampling distribution.

6. For a population of 17-year-old boys and 17-year-old girls, the means and standard deviations, respectively, of their subscapular skinfold thickness values are as follows: boys, 9.7 and 6.0; girls, 15.6 and 9.5. Simple random samples of 40 boys and 35 girls selected from the populations. What is the probability that the difference between sample means $\bar{x}_{girls} - \bar{x}_{boys}$ will be greater than 10?

7. We wish to estimate the average number of heartbeats per minute for a certain population. The average number of heartbeats per minute for a sample of 49 subjects was found to be 90. Assume that these 49 patients constitute a random sample, and that the population is normally distributed with a standard deviation of 10. Construct 90, 95, and 99 percent confidence intervals for the population mean, and state the practical and probabilistic interpretations of each. Explain why the three intervals that you construct are not of equal width. Indicate which of the three intervals you would prefer to use as an estimate of the population mean, and state the reason for your choice.

8. The maximal oxide diffusion rate in a sample of 15 asthmatic schoolchildren and 15 controls was reported as mean $\pm$ standard error of the mean. For asthmatic children, they reported 3.5 $\pm$ 0.4 nL/s (nanoliters per second) and for control subjects they reported 0.7 $\pm$ 0.1 nL/s. For each group, determine the following:

   (a) What was the sample standard deviation?

   (b) What is the 95% confidence interval for the mean maximal nitric oxide diffusion rate of the population?

   (c) What assumptions are necessary for the validity of the confidence interval you constructed?

   (d) What are the practical and probabilistic interpretations of the interval you constructed?

   (e) If you were to construct a 90% confidence interval for the population mean from the information given here, would the interval be wider or narrower than the 95% confidence interval? Explain your answer without actually constructing the interval.

   (f) If you were to construct a 99% confidence interval for the population mean from the information given here, would the interval be wider or narrower than the 95% confidence interval? Explain your answer without actually constructing the interval.

9. A study was performed to examine free fatty acid concentrations in 18 lean subjects and 11 obese subjects. The lean subjects had a mean level of 299 μEq/L with a standard error of the mean of 30, while the obese subjects had a mean of 744 μEq/L with a standard error of the mean of 62. Construct 90, 95, and 99 percent confidence intervals for the difference between population means. State the assumptions that make your

method valid. State the practical and probabilistic interpretations of each interval that you construct. Consider the variables under consideration and state what use you think researchers might make of your results.