<u>Assignment 3: Due to 4 pm, Oct 1st, 2017</u>

*** Please name your homework file as 'Assignment3_Your name.pdf' in pdf format and send it at biostat_sjtu@163.com , thanks for your cooperation.**

1. The duration of time from first exposure to HIV infection to AIDS diagnosis is called the incubation period. The incubation periods of a random sample of 7 HIV infected individuals is given below (in years):

12.0 10.5

9.5 6.3

13.5 12.5

7.2

a. Calculate the sample mean.

b. Calculate the sample median.

c. Calculate the sample standard deviation.

d. If the number 6.3 above were changed to 1.5, what would happen to the sample mean, median, and standard deviation? State whether each would increase, decrease, or remain the same.

e. Suppose instead of 7 individuals, we had 14 individuals. (we added 7 more randomly selected observations to the original 7)

12.0 10.5 5.2

9.5 6.3 13.1

13.5 12.5 10.7

7.2 14.9 6.5

8.1 7.9

Make an educated guess of whether the sample mean and sample standard deviation for the 14 observations would increase, decrease, or remain roughly the same compared to your answer in part (c) based on only 7 observations. Now actually calculate the sample mean standard deviation to see if you were right. How does your calculation compare to your educated guess?

Why do you think this is?

2. In a random survey of 3,015 boys age 11, the average height was 146 cm, and the standard deviation (SD) was 8 cm. A histogram suggested the heights were approximately normally distributed. Fill in the blanks.

   a. One boy was 170 cm tall. He was above average by __________ SDs.

   b. Another boy was 148 cm tall. He was above average by __________ SDs.

   c. A third boy was 1.5 SDs below the average height. He was __________ cm tall.

   d. If a boy was within 2.25 SDs of average height, the shortest he could have been is

   __________ cm and the tallest is __________ cm.

   e. Here are the heights of four boys: 150 cm, 130 cm, 165 cm, 140 cm. Which

   description from the list below best fits each of the boys (a description can be used

   more than once)? Justify you answer

   - Unusually short.
   - About average.
   - Unusually tall.

3. Assume blood-glucose levels in a population of adult women are normally distributed withmean 90 mg/dL and standard deviation 38 mg/dL.

   a. Suppose the "abnormal range" were defined to be glucose levels outside of 1 standard deviation of the mean (i.e., either at least 1 standard deviation above the mean, or at least 1 standard deviation below mean). Individuals with abnormal levels will be retested. What percentage of individuals would be called "abnormal" and need to be retested? What is the normal range of glucose levels in units of mg/dL?

   b. Suppose the abnormal range were defined to be glucose levels outside of 2 standard deviations of the mean. What percentage of individuals would now be called "abnormal"? What is the normal range of glucose levels (mg/dL)?

4. Suppose a random sample of 100 12-year-old boys were chosen and the heights of these 100 boys recorded. The sample mean height is 64 inches, and the sample standard deviation is 5 inches. You may assume heights of 12-year-old boys are normally distributed. Which interval below includes approximately 95% of the heights of 12-year-old boys?

   a. 63 to 65 inches.
   b. 39 to 89 inches.

c. 54 to 74 inches.
d. 59 to 69 inches.
e. Cannot be determined from the information given.
f. Can be determined from the information given, but none of the above choices is correct.

5. Cholesterol levels are measured on a random sample of 1,000 persons, and the sample standard deviation is calculated. Suppose a second survey were repeated in the same population, but the sample size tripled to 3,000. Then which of the following is true?

a. The new sample standard deviation would tend to be smaller than the first and approximately about one-third the size.

b. The new sample standard deviation would tend to be larger than the first and approximately about three times the size.

c. The new sample standard deviation would tend to be larger than the first, but we cannot approximate by how much.

d. None of the above is true because there is no reason to believe one standard deviation would tend to be larger than the other.

6. Suppose birth weights in a population are normally distributed with a mean of 109 oz and a standard deviation of 13 oz,

a. simulate (randomly take) 5 subjects (n=5) from the distribution of the birth weights (generate a normal distribution with mean of 109 and standard devialtion of 13), compute the mean of the sample, and then repeat the process 500 times in order to approximate the population mean and calculate the standard error. And then perform same simulations with n=20, 100 and 500. Display the boxplots for **the distribution of means of samples** with different sample sizes.

b. estimate the chance of obtaining a birth weight of 141 oz or heavier by taking 1000 random samples from the distribution of the birth weight above, each with 100 subjects. If the standard deviation increases to 26 oz, do estimation again by simulation.