

Name: \_\_\_\_\_

**University of Massachusetts: School of Public Health and Health Sciences  
BioEpi 540-Fall 2009**

**Exam 1**

Answer the questions by circling **all** correct answer or writing the answer on this paper. Include all work. Upon completing the exam, please email as an attachment your spreadsheet to: [bioep540@schoolph.umass.edu](mailto:bioep540@schoolph.umass.edu) . You may use books, calculators, computers and any notes, but you must do your own work.

**Question 1. (10 pts)** For a certain population, diastolic blood pressure has been evaluated to have a mean value of 75, a 10<sup>th</sup> percentile of 55, an 80<sup>th</sup> percentile of 90, and a range of 60.

**1A.** (5 pts) What can we say *for sure* about diastolic blood pressure in that population?

- A) 40% of the values are between 55 and 75
- B) 30% of the values are between 75 and 90
- C) 50% of the values are between 55 and 90
- D) The third quartile (Q3) cannot be greater than 90
- E) The maximum value is above 115
- F) Something else or not enough information

**1B.** (5 pts) Suppose a single subject is selected via simple random sampling from the population. What can we say for sure about X, the random variable representing the selected subject's diastolic blood pressure?

- A) The value of X is 75.
- B) The value of X is greater than 115.
- C) The value of X is in between 55 and 90.
- D) Nothing can be said about X, since it is a random variable.
- E) The expected value of X is 75.
- F) None of the above.

**Question 2. (25 pts)** The average saturated fat intake for each of 5 subjects in a population is given in Table 1.

Table 1. Average saturated fat intake in grams/day for  $N=5$  subjects.

ID	Sat Fat
A	3
B	6
C	16
D	10
E	10

a (15 pts). Calculate the population mean, median, and the population variance.

b. (5 pts). Suppose a simple random sample of  $n = 3$  subjects with replacement is selected. How many possible sample sequences could be selected?

c. ( 5 pts). If the sample of size  $n = 3$  was selected without replacement, how many possible sample sets could be selected.

**Question 3 (75 pts).** A study is to be conducted using simple random sampling without replacement, where  $n = 3$ . There is interest in estimating the median saturated fat intake in the population based on results from a single sample set. Three candidate estimators are:

- i. The sample median
- ii. The sample mean
- iii. The average of the sample minimum and maximum value.

You have been asked to recommend how to estimate the population median. You should use a spreadsheet to help answer these questions. Upon completing these questions, Email your spreadsheet as an attachment to: [bioep540@schoolph.umass.edu](mailto:bioep540@schoolph.umass.edu) .

- a. (10 pts) Using a spreadsheet, list all possible simple random samples without replacement. (Do not list results here).
- b. (5 pts) One possible sample set includes subjects A, B, and C. What is the probability that this sample will be selected?
- c. (10 pts) In your spreadsheet, evaluate each of the estimators for each of the possible samples. (Do not list results here.)
- d. (10 pts). Using your spreadsheet, determine the expected value of each of the estimators, and list these results below.

E(sample median)=

E(sample mean)=

E( (min+max)/2) =

- e. (10 pts). When estimating the median, which estimators are biased, and which estimators are unbiased. Which estimator has the smallest bias?

- f. (10 pts). Evaluate the variance of each of the estimators.

var(sample median)=

var(sample mean)=

var( (min+max)/2) =

g. (10 pts). Evaluate the mean squared error of each estimator.

MSE(sample median)=

MSE(sample mean)=

MSE( (min+max)/2) =

h. (10 pts). Based on the results that you have evaluated in questions 3e-3g, which estimators would you recommend using to estimate the population median? Why would you recommend it over the others?