

Name: _____ SOLUTION _____

**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 . 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 10th percentile of 55, an 80th 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.

$$\begin{aligned}\mu &= \frac{1}{N} \sum_{s=1}^N y_s \\ &= \frac{1}{5} (3 + 6 + 16 + 10 + 10) \\ &= 9\end{aligned}$$

Median= 10

$$\begin{aligned}\sigma^2 &= \frac{1}{N} \sum_{s=1}^N (y_s - \mu)^2 \\ \text{Population Variance} &= \frac{1}{5} \left((-6)^2 + (-3)^2 + (7)^2 + (1)^2 + (1)^2 \right) \\ &= 19.2\end{aligned}$$

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

$$N^n = (5)^3 = 125$$

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

$$\begin{aligned}\binom{N}{n} &= \frac{N!}{(N-n)!n!} \\ &= \frac{5!}{3!2!} = \frac{5(4)}{2} = 10\end{aligned}$$

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 .

- 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?

$$\text{Prob}(\text{sample } \{A,B,C\}) = 1/10$$

- 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.

$$E(\text{sample median}) = 8.8$$

$$E(\text{sample mean}) = 9$$

$$E((\text{min} + \text{max})/2) = 9.1$$

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

All estimators are biased, since the population median is 10. The estimator that is the average of the minimum and maximum has the smallest bias.

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

$$\text{var}(\text{sample median}) = 3.36$$

$$\text{var}(\text{sample mean}) = 3.20$$

$$\text{var}((\text{min} + \text{max})/2) = 4.44$$

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

MSE(sample median)=4.80

MSE(sample mean)=4.2

MSE((min+max)/2) = 5.25

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?

I would recommend using the sample mean to estimate the sample median since the MSE is smallest for the sample mean.

A	B	C	D	E											
3	6	16	10	10											
Subj 1	Subj 2	Subj 3	Resp1	Resp2	Resp3	Sample	Estimate Median	Estimate Mean	Estimate (min+max)/2	Var Median	Var Mean	Var (min+max)/2	MSE Median	MSE Mean	MSE (min+max)/2
A	B	C	3	6	16	1	6	8.33	9.5	7.84	0.44	0.16	16	2.778	0.25
A	B	D	3	6	10	2	6	6.33	6.5	7.84	7.11	6.76	16	13.444	12.25
A	B	E	3	6	10	3	6	6.33	6.5	7.84	7.11	6.76	16	13.444	12.25
A	C	D	3	16	10	4	10	9.67	9.5	1.44	0.44	0.16	0	0.111	0.25
A	C	E	3	16	10	5	10	9.67	9.5	1.44	0.44	0.16	0	0.111	0.25
A	D	E	3	10	10	6	10	7.67	6.5	1.44	1.78	6.76	0	5.444	12.25
B	C	D	6	16	10	7	10	10.67	11	1.44	2.78	3.61	0	0.444	1
B	C	E	6	16	10	8	10	10.67	11	1.44	2.78	3.61	0	0.444	1
B	D	E	6	10	10	9	10	8.67	8	1.44	0.11	1.21	0	1.778	4
C	D	E	16	10	10	10	10	12.00	13	1.44	9.00	15.21	0	4.000	9
						Average	8.800	9	9.100	3.360	3.200	4.440	4.800	4.200	5.250

A	B	C	D	E			Estimate	Estimate	Estimate	Var	Var	Var	MSE	MSE					
3	6	16	10	10	Subj 1	Subj 2	Subj 3	Resp1	Resp2	Resp3	Sample	Median	Mean	(min+max)/2	Median	Mean	(min+max)/2	Median	Mean
A	B	C	3	6	16	1	6	8.33	9.5	7.84	0.4444444	0.16	16	2.7777778					
A	B	D	3	6	10	2	6	6.33	6.5	7.84	7.1111111	6.76	16	13.4444444					
A	B	E	3	6	10	3	6	6.33	6.5	7.84	7.1111111	6.76	16	13.4444444					
A	C	D	3	16	10	4	10	9.67	9.5	1.44	0.4444444	0.16	0	0.1111111					
A	C	E	3	16	10	5	10	9.67	9.5	1.44	0.4444444	0.16	0	0.1111111					
A	D	E	3	10	10	6	10	7.67	6.5	1.44	1.7777778	6.76	0	5.4444444					
B	C	D	6	16	10	7	10	10.67	11	1.44	2.7777778	3.61	0	0.4444444					
B	C	E	6	16	10	8	10	10.67	11	1.44	2.7777778	3.61	0	0.4444444					
B	D	E	6	10	10	9	10	8.67	8	1.44	0.1111111	1.21	0	1.7777778					
C	D	E	16	10	10	10	10	12.00	13	1.44	9	15.21	0	4					
							Average	8.800	9	9.100	3.360	3.200	4.440	4.800	4.200				