BIOSTATS 540 - Introductory Biostatistics
Fall 2016
Examination 1 (Unit 1 – Summarizing Data)
Due: Monday September 26, 2016
Last Date for Submission with Credit: Monday October 3, 2016

Rules:
This is an "open book" "take-home" exam. You are welcome to use any reference materials you wish. You   are welcome to use the computer as you wish, too. However, you MUST work this exam by yourself and you may not consult with anyone.

Instructions and Checklist

___1. Start each problem on a new page.
___2. Write your name on every page
___3. Make a copy of your exam for safekeeping (sometimes a mailed exam is lost!)
___4. Submit a completed signature page (See next page).

How to submit your exam:

Worcester "in-class" Section:
___1. Bring your completed exam to class on **Monday September 26, 2016** being sure that you have
     made a copy for safe keeping; OR
___2. Mail your completed exam to me with post-mark **September 26, 2016** to my address below.

Blackboard Learn Online Section:
___1. Upload your completed exam to the ASSIGNMENT tab no later than 11:59 pm on **Monday**
     September 26, 2016. This must be a single pdf ;   OR
___2. Mail your completed exam to me with post-mark **September 26, 2016** to my address below.

Address and Telephone Number for Mailing

Carol Bigelow
Biostatistics & Epidemiology/402 Arnold House
University of Massachusetts/Amherst
715 North Pleasant Street
Amherst, MA 01003-9304
Tel. 413-545-1319

BIOSTATS 540 - Introductory Biostatistics
Fall 2016
Examination 1 (Unit 1 – Summarizing Data)
DUE: 11:59 PM Monday September 26, 2016

Signature

This is to confirm that in completing this exam, I worked independently and did not consult with anyone.

Signature: _____

Printed Name: _____

Date: _____

1. (20 points total)

Consumer Reports magazine reported the following data on the number of calories in a hot dog for each of a sample of 17 brands of meat hot dogs

173	191	182	190	172	147	146	139	175
136	179	153	107	195	135	140	138	

1a. (5 points)
By any approach you like (by hand, using Excel, or using a statistical software of your choosing), calculate the **sample mean, sample variance,** and **sample standard deviation.** Show your work ("cut and paste" screen capture is fine).

1b. (5 points)
By any approach you like (by hand, using Excel, or using a statistical software of your choosing), calculate the **standard error** of the sample mean. Show your work ("cut and paste" screen capture is fine).

1c. (5 points)
In 1-2 sentences at most, explain the distinction between the standard deviation and the standard error.

1d. (5 points)
By any approach you like (by hand, using Excel, or using a statistical software of your choosing), calculate the **five-point summary.** Show your work ("cut and paste" screen capture is fine).

2. (20 points total)

2a. (4 points) – Multiple Choice (Choose ONE)
If a distribution is skewed to the right,

    __A. The mean is less than the median
    __B. The mean and the median are equal
    __C. The mean is greater than the median

2b. (4 points) – Multiple Choice (Choose ONE)
What percent (%) of the observations in a distribution lie between the first quartile and the third quartile?

    __A. 25%
    __B. 50%
    __C. 75%

2c. (4 points) – Multiple Choice (Choose ONE)
What are all the values that a standard deviation (s) can possibly take?

    __A. $0 \leq s$
    __B. $0 \leq s \leq 1$
    __C. $-1 \leq s \leq 1$

2d. (4 points) – Multiple Choice (Choose ONE)
You have data on the weights in grams of 5 baby pythons. The mean weight is 31.8 and the standard deviation of the weights is 2.39. The correct units for the standard deviation are

    __A. no units; it's just a number
    __B. grams
    __C. grams squared

2e. (4 points) – Multiple Choice (Choose ONE)
Which of the following is *least affected* if an extreme high outlier is added to your data

    __A. The median
    __B. The mean
    __C. The standard deviation

3. (20 points total)

The table below shows the distribution of dietary vitamin-A intake as reported by 14 students who filled out a dietary questionnaire in class. The total intake is a combination of intake from individual food items and from vitamin pills. The units are in IU/100 ((international Units/100).

Student Number	Intake (IU/100)	Student Number	Intake (IU/100)
1	31.1	8	48.1
2	21.5	9	24.4
3	74.7	10	13.4
4	95.5	11	37.1
5	19.4	12	21.3
6	64.8	13	78.5
7	108.7	14	17.7

3a. (5 points)
By any approach you like (by hand, using Excel, or using a statistical software of your choosing), compute the **sample mean** and **sample median.** Show your work ("cut and paste" screen capture is fine).

3b. (5 points)
By any approach you like (by hand, using Excel, or using a statistical software of your choosing), compute the **standard deviation** and **coefficient of variation.**

3c. (5 points)
Suppose the data are expressed in IU rather than IU/100. What are the **sample mean, standard deviation,** and **coefficient of variation** in the new units?

3d. (5 points)
Which do you think is a more appropriate measure of location for this data set, the **mean** or the **median**? In 1 sentence at most, explain your reasoning.

4. (20 points total)

A zoologist measured tail length in 86 individuals, all in the one-year age group, of the deer mouse *Peromyscus*. The mean length was 60.43 mm and the standard deviation was 3.06 mm.

4a. (4 points)
By hand, calculate the standard error of the mean. Show your work.

The following pertains to questions #4b and #4c:
Suppose the zoologist were to measure 500 additional animals from the same population.

4b. (5 points)
What would you predict would be the **standard deviation** of the 500 new measurements?

4c. (5 points)
What would you predict would be the **standard error of the mean** of the 500 new measurements?

The following pertains to questions #4d, #4e, and #4f:
For each of the following, decide whether the description fits the standard deviation (s) or the standard error of the mean (sem).

4d. (2 points)
This quantity is a measure of the accuracy of the sample mean as an estimate of the population mean

4e (2 points)
This quantity tends to stay the same as the sample size goes up.

4f. (2 points)
This quantity tends to go down as the sample size goes up.

5. (20 points total)

The following is a tabulation of some imaginary data.

Table -
Number of Cases of Acute Gastroenteritis Occurring in an Imaginary City, 1980-2010

Year	Number of Cases
1980	400
1985	600
1990	800
1995	900
2000	1,000
2005	1,100
2010	1,200

5a. (10 points)
State the facts shown in this table.

5b. (10 points)
Just give this a try!
I will give credit to any ideas you have that are reasonable! Spend some time studying this table, working with the numbers in any way you like.   In 1 or 2 paragraphs (it doesn't have to be long), **what do you conclude?** Show your work.