

**PubHlth 540 - Introductory Biostatistics**  
**Fall 2008**  
**Examination I**  
**DUE: Post mark Monday October 27, 2008**

**Before you begin:**

*This is a “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 photo-copy of your exam for safekeeping prior to submission
- \_\_\_ 4. Complete the signature page

**How to submit your exam:**

<b>Online Section</b>	<b>Worcester Section</b>
(1) By mail (address and telephone number below) and post- marked no later than 11:59 PM Monday October 27, 2008  <p style="text-align: center;">OR</p> (2) By upload to EXAM I turn in link no later than 11:59 PM Monday October 27, 2008	(1) By mail (address and telephone number below) and post- marked no later than 11:59 PM Monday October 27, 2008  <p style="text-align: center;">OR</p> (2) Bring your exam to class Monday October 27, 2008
<i>Faxed exams are NOT permitted - sorry</i>	<i>Faxed exams are NOT permitted – sorry.</i>

**Address and telephone number for mailing**

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

**PubHlth 540 - Introductory Biostatistics**  
**Fall 2008**  
**Examination I**  
**DUE: Post mark Monday October 27, 2008**

**Signature**

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

Signataure: \_\_\_\_\_

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

Date: \_\_\_\_\_

**1. (10 points total)**

The Honolulu Heart Study, conducted in the 1970's, measured lots of things on lots of study participants. The following are 10 of the study variables. For each, please tell me two things: (1) whether the variable measured was qualitative or quantitative. If qualitative, indicate (2) whether the variable measured was nominal or ordinal. If quantitative, indicate (2) whether the variable measured was discrete or continuous

		Qualitative or Quantitative	If qualitative: nominal or ordinal If quantitative: discrete or continuous
<b>1 point</b>	<u>Education level:</u> 1=none 2=primary 3=intermediate 4=senior high 5=technical school 6=university		
<b>1 point</b>	<u>Weight:</u> kilograms		
<b>1 point</b>	<u>Height:</u> centimeters		
<b>1 point</b>	<u>Age:</u> Years		
<b>1 point</b>	<u>Smoking status:</u> 0=no 1=yes		
<b>1 point</b>	<u>Physical activity:</u> 1=mostly sitting 2= moderate 3=active		
<b>1 point</b>	<u>Blood glucose:</u> Milligrams percent		
<b>1 point</b>	<u>Serum cholesterol:</u> Milligrams/deciliter		
<b>1 point</b>	<u>Systolic blood pressure</u> Millimeters mercury		
<b>1 point</b>	<u>Ponderal index:</u> $\text{Height}^3 \sqrt{\text{weight}}$		

**2. (10 points total)**

The following is a table showing the frequency and cumulative frequency distribution of the survival times of 347 patients diagnosed with cancer.

Survival Time (years)	Frequency	Cumulative Frequency
<1	62	62
1-2	45	107
2-3	38	145
3-4	28	173
4-5	25	198
5-6	10	208
6-7	14	222
7-8	11	233
8-9	9	242
9-10	8	250
10-11	8	258
11-12	8	266
12-13	9	275
13-14	5	280
14-15	2	282
15-16	3	285
16-17	4	289
17-18	7	296
18-19	1	297
19-20	3	300
At least 20	47	347
<b>Total</b>	<b>347</b>	

- (a) **(3 points)** What is the **modal** survival time?
- (b) **(3 points)** Estimate the **median** survival time.
- (c) **(2 points)** Estimate the smallest value of the **mean** survival time.
- (d) **(2 points)** In your opinion, are these data symmetric, positively skewed, or negatively skewed?  
Explain your answer

**3. (5 points total)**

Consider the relationship between the standard deviation and the standard error. Suppose it is known that the standard deviation is 3. How large a sample  $n$  should be taken for the standard error of the mean to have a value of 0.5?

**4. (10 points total)**

Consider the following cross-tabulation of 100 individuals by their smoking status and systolic blood pressure. Shown are counts of individuals. For example, the entry “10” in the first row says that there are 10 non-smokers in this sample with systolic blood pressure between 90 and 109 mm Hg.

Systolic blood pressure, mm Hg	Non-Smokers	Smokers
90-109	10	5
110-129	24	15
130-149	18	10
150-169	9	3
170-189	2	2
190-209	0	3

Consider this distribution as a population and, as such, a universe of possibilities from which simple probabilities can be computed.

Define two events “A” and “B” as follows:

A = smoker

B = systolic blood pressure of 170 or greater.

- (a) **(2 points)** Find Probability [ A ]
- (b) **(2 points)** Find Probability [ B].
- (c) **(6 points)** Are A and B independent? Explain.

**5. (15 points total)**

It's not just an exercise in algebra that there is sometimes interest in re-location and re-scaling. One example of re-location and re-scaling is the conversion of temperature information expressed in degrees centigrade to degrees fahrenheit. A second example, closer to public health, is the use of international units IU. Weight reported in kilograms or pounds and height reported in centimeters or inches are other examples.

Suppose that the sample mean and sample standard deviation of  $n=50$  values of a random variable  $X$  are  $\bar{X} = 100$  and  $s = 15$ . (note – this is close to the values of the population mean and standard deviation of IQ).

- (a) **(5 points)** A constant  $b=5$  is added to each observation. What are the new values of the

**1 point.** Sample mean = \_\_\_\_\_,  
**1 point** Sample standard deviation = \_\_\_\_\_, and  
**3 points** Sample variance = \_\_\_\_\_

- (b) **(5 points)** Next, consider instead that each observation is multiplied by a constant  $a=5$ . What are the new values of the

**1 point.** Sample mean = \_\_\_\_\_,  
**1 point** Sample standard deviation = \_\_\_\_\_, and  
**3 points** Sample variance = \_\_\_\_\_

- (c) **(5 points)** Now suppose that the sample mean and sample standard deviation of 100 measurements of temperature in centigrade are  $\bar{X}$  and  $s$ . What is the expression for the same sample mean and sample standard deviation when the temperature values are converted to fahrenheit?

*Hint: [Degrees fahrenheit] = (1.8)[Degrees centigrade] + 32*

**6. (10 points total)**

- (a) **(5 points)** The following is a relative frequency table summarizing the score earned on a PubHlth 540 exam. What is the average score?

<b>Exam I Score</b>	<b>Relative Frequency</b>
100	.40
88	.30
81	.20
76	.10

- (b) **(5 points)** As the Fall 2008 election draws near, we find ourselves on the brink of a disastrous recession. During recessions, average earnings typically go up. Explain how this can be.

**7. (10 points total)**

Consider the following data on a new screening test for diabetes

	Diabetes	NO diabetes	Total
Test +	59	48	107
Test -	11	462	473
	70	510	580

(a) **(4 points)** Using the information in the table, compute the following:

- 1 point.** Sensitivity = \_\_\_\_\_,  
**1 point** Specificity = \_\_\_\_\_  
**1 point** Predictive value positive = \_\_\_\_\_  
**1 point** Predictive value negative = \_\_\_\_\_

(b) **(4 points)** Using the values of sensitivity and specificity from part “a” and *not using* your values of predictive value positive from part “a”, use your understanding of *Bayes rule* to complete the following table

Prevalence Diabetes	Predictive Value + Test
.01	
.05	
.10	
.20	
.30	
.40	
.50	
.60	
.70	
.80	
.90	
.95	
.99	

(c) **(2 points)** True or False.

- 1 point.** \_\_\_\_ The **sensitivity** of a diagnostic test is independent of the probability of disease
- 1 point** \_\_\_\_ The **predictive value** of a diagnostic test is independent of the probability of disease.

**8. (10 points total)**

A scientist has just been funded by the National Institutes of Health to conduct a very expensive study. As it happens, the budget has been cut. Therefore, the choice of laboratory for analysis of sample data is being re-visited.

Suppose the choice is between laboratory “A” and “B”. Laboratory “A” has a standard deviation of one milligram and each sample analysis costs \$3. Laboratory “B” has a standard deviation of two milligrams but each sample analysis costs \$1.

All other things being equal, is the principal investigator better off spending \$9 on laboratory “A” or spending \$9 on laboratory “B”? *Hint – This question is asking you to show me your understanding of standard error.*

**9. (10 points total)**

Let “A” and “B” denote two independent genetic traits and suppose that the probability that a randomly selected individual will exhibit trait “A” is  $\frac{1}{2}$ . Suppose also that the probability that a randomly selected individual will exhibit trait “B” is  $\frac{3}{4}$ .

What is the probability that a randomly selected individual will exhibit

- (a) **2 points.** Both traits? = \_\_\_\_\_,
- (b) **2 points.** Neither? = \_\_\_\_\_,
- (c) **2 points.** Trait A but not trait B? = \_\_\_\_\_,
- (d) **2 points.** Trait B but not trait A? = \_\_\_\_\_,
- (e) **2 points.** Exactly one trait? = \_\_\_\_\_,

**10. (10 points total)**

The following are data from a **case-control** study that investigated the relationship between helmet use and facial injury following a bicycle accident.

Helmet Worn		Facial Injury	
		Yes	No
No	182	236	
Yes	30	83	

- (a) **(2 points)** Is this study prospective or retrospective?.
- (b) **(2 points)** What are the odds of a facial injury among persons who did not wear a helmet?
- (c) **(2 points)** What are the relative odds (odds ratio) of facial injury for persons who did not wear a helmet compared to those who did wear a helmet?
- (d) **(2 points)** Is the calculation of the relative risk meaningful in this setting? Explain.
- (e) **(2 points)** Considering your answers to parts “a” through “d”, in your opinion, does wearing a helmet reduce the risk of sustaining a facial injury in a bicycle accident? Explain.