

**PubHlth 540 - Introductory Biostatistics
Fall 2009**

Examination I

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:

| Online Section | Worcester Section |
|---|--|
| (1) By mail (address and telephone number below) and post- marked no later than 11:59 PM Monday October 26, 2009 <p style="text-align: center;">OR</p> (2) By upload to EXAM I turn in link no later than 11:59 PM Monday October 26, 2009 | (1) By mail (address and telephone number below) and post- marked no later than 11:59 PM Monday October 26, 2009 <p style="text-align: center;">OR</p> (2) Bring your exam to class Monday October 26, 2009 |
| <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 2009
Examination I
DUE: 11:59 PM Monday October 26, 2009

Signature

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

Signature: _____

Printed Name: _____

Date: _____

1. (10 points total)

Identify the scale (eg – discrete/continuous and nominal/ordinal or interval/ratio)for the following variables.

- 1a (1 point). Calories consumed during the day
- 1b (1 point) Marital status
- 1c (1 point) Perceived health status reported as “poor”, “fair”, “good” or “excellent”
- 1d (1 point) Blood type
- 1e (1 point) IQ score

For the following health interview survey items, propose how the variable should be defined for analysis purposes. Specifically, suggest a variable name and coding scheme for its measurement. **(Eg for age: variable name: AGE coding scheme: years)**

- 1f (1 point) Weight
- 1g (1 point) Height
- 1h (1 point) Family Income
- 1i (1 point) Unemployment
- 1j (1 point) Number of stays in a mental hospital

2. (10 points total)

The following table summarizes data obtained from a survey of 4390 persons regarding their awareness of suicide. The question asked of each study participant was: *“How many people have you ever known who were victims of suicide?”*

| Number of Victims | Frequency |
|-------------------|-------------|
| 0 | 3944 |
| 1 | 279 |
| 2 | 97 |
| 3 | 40 |
| 4 or more | 30 |
| Total | 4390 |

2a. (2 points).

Calculate the sample mean number of victims reported. In doing your calculations, give a score of 4.5 to the response “4 or more”.

2b. (2 points).

Calculate the sample median number of victims reported.

2c. (2 points)

If 1744 observations shift from a value of 0 to a value of “4 or more”, how does the sample mean change? That is, what is the new value of the sample mean?

2d. (2 points)

If 1744 observations shift from a value of 0 to a value of “4 or more”, how does the sample median change? That is, what is the new value of the sample median?

2e. (2 points)

Consider a study of n=60 subjects in which subjects are asked “How many serious motor vehicle accidents have you had in the past 5 years? You are asked to summarize the responses obtained and you need to choose between reporting the sample mean or the sample median. Which statistic would you choose to report and why?

3. (10 points total)

3a. (2 points).

True or False:

The sample mean, sample median and sample mode can never be all the same.

3b. (2 points).

True or False:

The sample mean is always one of the data points.

3c. (2 points).

True or False:

When the sample size “n” is odd, the sample median is one of the data points.

3d. (2 points).

According to a newspaper article, the mean wage for a sports player of a particular sport in the U.K. was 676,000 pounds. Suppose it is known that the distribution of wages is skewed to the right. True or False: The median salary is greater than 676,000 pounds.

3e. (2 points)

According to a recent National Center for Health Statistics Survey, among males aged 25-34 years: 2% have heights equal to 64 inches or less, 8% have heights equal to 66 inches or less, 27% have heights equal to 68 inches or less, 39% have heights equal to 69 inches or less, 54% have heights equal to 70 inches or less, 68% have heights equal to 71 inches or less, 80% have heights equal to 72 inches or less, 93% have heights equal to 74 inches or less, and 98% have heights equal to 76 inches or less. Which category of height has the median height? Explain your answer.

4. (10 points total)

A census survey taken by the U.S. Bureau of the Census reports that median earnings in the past 12 months was \$32,168 for women and \$41,975 for men. It also reported that mean earnings in the past 12 months was \$39,890 for women and \$56,724 for men.

4a. (5 points).

Does this data suggest that the distribution of income for each gender is symmetric, positive skewed (right skewed) or negative skewed (left skewed)? Explain your reasoning.

4b. (5 points)

The results were obtained from a survey of 73.8 million women and 83.4 million men. Calculate the overall sample mean income.

5. (10 points total)

A local arts and crafts store has found, after many years of experience, 20% of people who enter the store actually buy something.

Suppose 3 people enter the store.

5a. (5 points).

How many possible outcomes are possible for the sale or non-sale of items to the three potential customers? *Hint: In developing your answer, you are actually specifying the sample space for the outcome of sales and non-sales to the three potential customers.*

5b. (5 points).

What is the probability of at least one sale?

6. (10 points)

The following table summarizes results on an ELISA test used to aid in diagnosing Anthrax. These results are based on tests performed on 12 Anthrax cases, and 18 persons who did not have Anthrax. Use this table to answer the following questions.

ELISA for poly-D-glutamic acid capsule

| | Anthrax Case | No Anthrax | Total |
|----------------------|---------------------|-------------------|--------------|
| Test Positive | 11 | 2 | 13 |
| Test Negative | 1 | 16 | 17 |
| Total | 12 | 18 | 30 |

- 6a. (2 points)
What is the sensitivity of the test?

- 6b. (2 points)
What is the specificity of the test?

- 6c. (3 points)
Anthrax is relatively rare in the US population. Assume that the prevalence of Anthrax is 1 in 1,000,000. Now, suppose a simple random sample of 1,000,000 people are tested for Anthrax. Among the 1,000,000 people tested, what proportion would you expect to have positive test results?

- 6d. (3 points)
How many false positive results would you expect as a result of the 1,000,000 tests?

7. (10 points total)

Suppose you take your temperature once a week for twelve weeks and the values are 98.2, 99.0, 98.5, 99.1, 98.6, 97.9, 98.2, 99.3, 98.1, 98.5, 99.0, and 98.8.

7a. (5 points)

Compute the sample range, sample variance, and sample standard deviation.

7b. (5 points)

What proportion of the twelve observations are within ± 0.68 standard deviations of the mean?

Hint – In developing your answer, you will need to compute the sample mean.

8. (10 points total)

8a. (5 points)

Perhaps you've heard the saying "if you observe for long enough, a monkey will eventually write a Shakespeare play just by chance". **Hah!** Consider the setting of putting a typewriter in the hands of a monkey. Assume the typewriter has 50 keys and the monkey types key after key at random. What is the probability that the first seven keys that the monkey types is the seven character title *macbeth*?

8b. (5 points)

Consider a jury trial setting in which the probability that a defendant is convicted, given he or she is actually guilty, is 0.95 and that the probability that a defendant is acquitted, given that he or she is actually innocent is 0.95. Suppose further that 90% of all defendants are actually guilty. If it is known only that a defendant is convicted, what is the probability that he or she is actually innocent?

9. (10 points total)

9a. (5 points)

Suppose the scores on a difficult exam have a sample mean of 67 and a standard deviation of 20. Suppose the instructor decides to boost each student's score by 10 points; that is, the instructor adds 10 points to each student's score. What are the values of the new sample mean and sample standard deviation?

9b. (5 points)

Next, suppose that the sample mean income for a group of workers is \$59,000 with a standard deviation of \$15,000. You are hired to report this information at a meeting in the U.K. If one British pound equals \$2.25, what is the value of the sample mean and sample standard deviation in units of British pounds?

10. (10 points total)

The following are hypothetical data from a study of the relationship between cigarette smoking and the risk of low birth weight. Suppose preliminary analyses included the classification of infants into categories according to the number of cigarettes smoked per day by the mother. In particular, suppose the probabilities of the different smoking levels are the following:

| Cigarettes per day | Probability |
|--------------------|---------------|
| Zero | 0.9082 |
| 1-5 | 0.0855 |
| 6-20 | 0.0059 |
| > 20 | 0.0004 |
| Total | 1.0000 |

Next, suppose the following *conditional* probabilities are known.

| Condition: Mother smokes | Conditional Probability of low birth weight infant |
|-----------------------------|---|
| Zero cigarettes/day | 0.035 |
| 1-5 | 0.379 |
| 6-20 | 0.813 |
| > 20 | 0.540 |

10a. (5 points)

Calculate the probability of a low birth weight infant.

10b. (5 points)

Are the events “>20 cigarettes per day” and “low birth weight infant” independent? Explain your reasoning using the rules of probability developed in class.