BIOSTATS 540 - Introductory Biostatistics
Fall 2015
Examination 2
Due: Monday October 19, 2015
Last Date for Submission with Credit: Monday October 26, 2015

Ground 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.

How to submit your exam:

Worcester “in-class” Section:
__1. Bring your completed exam to class on Monday October 19, 2015, being sure that you have made a copy for safe keeping; OR
__2. Mail your completed exam to me with post-mark October 19, 2015 to my address below.

Online Section:
__1. Upload your completed exam to the ASSIGNMENT tab no later than 11:59 pm on Monday October 19, 2015. This must be a single pdf and be named using the convention lastname_exam1.pdf. OR
__2. Mail your completed exam to me with post-mark October 19, 2015 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
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)

1a. (2 points)
Various claims are often made for surveys. Is the following claim correct? If not, why? 
“Stopping students on their way out of the cafeteria is a good way to sample if we want to know about the quality of the food there.”

1b. (2 points)
Various claims are often made for surveys. Is the following claim correct? If not, why? 
“A poll taken of a statistics support website garnered, 12,357 responses. The majority said they enjoy doing statistics homework. With a sample size that large, we can be pretty sure that most statistics students feel this way too.”

1c. (2 points)
Various claims are often made for surveys. Is the following claim correct? If not, why? 
“The true percentage of all statistics students who enjoy the homework is called a ‘population statistic’”

1d. (2 points)
A survey of a random sample of 300 passengers on a flight from San Francisco to Tokyo was undertaken. The sampling method was the following: pick every 10th passenger as people board the plane. What is the name of this sampling method?

1e. (2 points)
A survey of a random sample of 300 passengers on a flight from San Francisco to Tokyo was undertaken. The sampling method was the following: from the boarding list, randomly choose 5 people flying first class and 25 of the other passengers. What is the name of this sampling method?
2. (10 points total)

A question posted on a gaming website asked visitors to this site, “Do you have an active social life outside the internet?” 22% of the 55,581 respondents said “No” or “Not really, most of my personal contact is through social media.”

Identify the following items of this survey research. If you feel an item is not identifiable then report your answer for that item as “not identifiable from the information given”.

2a. (1 point)
The population.

2b. (1 point)
The population parameter of interest.

2c. (1 point)
The sampling frame.

2d. (1 point)
The sample.

2e. (2 points)
The sampling method, including whether or not randomization was employed.

2f (2 points)
Who (if anyone) was left out of the study.

2g (2 points)
Any potential sources of bias and any problems you see in generalizing to the population of interest (this is your answer to question #2a).
3. (15 points total)

Suppose that, every day, you drive through the intersection at Pleasant and Amity, Amherst, MA. As you might imagine, the light itself is governed by a timer. Suppose that your arrival at this intersection is completely random and that all arrival times are independent.

Finally, suppose that, upon arrival at the intersection, the probability of a green light is 0.35, a yellow light is 0.04, and a red light is 0.61.

3a. (5 points)
What is the probability that you arrive to a red light on both Monday and Tuesday?

3b. (5 points)
What is the probability that you don’t encounter a red light until Wednesday?

3c. (5 points)
What is the probability that you will have to stop for a red light at least once during the week? *Hint: Think “complement”*
4. (15 points total)

An airline offers discounted “advance-purchase” fares to customers who buy tickets more than 30 days before travel and charges “regular” fares for tickets purchased during those last 30 days. The company has noticed that 60% of its customers take advantage of the advance-purchase fares. The “no-show” rate among people who paid regular fares is 30%, but only 5% of customers with advance-purchase tickets are now-shows.

4a. (5 points)
What percent of all ticket holders are “no-shows”?

4b. (5 points)
What is the probability that a customer who didn’t show had an advance-purchase ticket?

4c. (5 points)
Is being a “no-show” independent of the type of ticket a passenger holds? Explain.
5. (10 points total)

Lie detectors are controversial instruments, barred from use as evidence in many courts. Nonetheless, many employers use lie detector screening as part of their hiring process in the hope that the can avoid hiring people who might be dishonest. There has been some research, but no agreement, about the reliability of polygraph tests. Based on this research, suppose that a polygraph can detect 65% of lies, but incorrectly identifies 15% of true statements as lies.

A certain company believes that 95% of its job applicants are trustworthy. The company gives everyone a polygraph test, asking: “have you ever stolen anything from your place of work?” Naturally, all the applicants answer “No” but the polygraph identifies some of those answers as lies, making person ineligible for a job.

5a. (5 points)
What is the probability that a job applicant rejected under suspicion of dishonesty was actually trustworthy?

5b. (5 points)
In 1-2 sentences, in your opinion, what do these values suggest about using lie detector tests to identify dishonest applicants?
6. (10 points total)

The magazine BusinessWeek ran an online poll on their website and asked the readers the following question: “Do you think Google is too powerful?” Poll participants were offered three buttons as response: “yes”, “no”, or “not sure”.

In all 1336 (35.9%) said “yes”, 2051 (55.1%) said “no” and 335 (9.0%) said “not sure”

6a. (5 points)
What is the sample size for this poll?

6b. (2 points).
The magazine BusinessWeek was careful to note that: “This was not a scientific poll”. Drawing upon your understanding of probability versus non-probability sampling plans, in your opinion, what was meant by this statement?

6c. (3 points).
Do you think this poll might have given unreliable information? In 1-2 sentences, explain your reasoning.
7. (20 points total)

A certain research company analyzed the nutritional content of 77 brands of packaged breakfast cereals. For each cereal, one of the measurements made was the fat content (in grams) per serving. Call this random variable X.

The following is the probability model for X = fat content (in grams) per serving.

<table>
<thead>
<tr>
<th>X</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.35</td>
</tr>
<tr>
<td>1</td>
<td>0.39</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
</tr>
<tr>
<td>3</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>5</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Consider the following two events:

A = \{ fat content is 4 grams or greater \}
B = \{ fat content is 5 grams or less \}

7a. (5 points)
What outcomes make up the event A? What is Probability [ A ]?

7b. (5 points)
What outcomes make up the event B? What is Probability [ B ]?

7c. (5 points)
What outcomes make up the event “A or B”? What is Probability [ A or B ]?

7d. (5 points)
Why is Probability [A or B] ≠ Probability[A] + Probability[B]?
8. (10 points total)

8a. (2 points) Fill in the blank. Choose the ONE BEST answer.
A simple random sample is a sample chosen in such a way that every unit in the population has a(n) ________ chance of being selected into the sample.

   __A.  equal  
   __B.  unequal  
   __C.  known 

8b. (2 points) Multiple Choice. Choose ONE.
In a random digit table, approximately what percent of the digits are 9 or 2?

   __A.  20%  
   __B.  10%  
   __C.  unknown 

8c. (2 points) TRUE or FALSE.
Sampling with replacement from a very large population gives virtually the same result as sampling without replacement.

8d. (2 points) Complete the sentence by choosing the ONE BEST last word
In a stratified random sample, the selection probability for each element within a stratum is __.

   __A.  equal  
   __B.  unequal  
   __C.  known 

8e. (2 points) Fill in the blank. Choose the ONE BEST answer.
A probability sample is a sample chosen in such a way that each possible sample has a (n) ______ chance of being selected.

   __A.  equal  
   __B.  unequal  
   __C.  known  
   __D.  unknown
A case-control study was conducted to investigate the association of consuming chili peppers and gastric cancer. **Cases (disease)** were 21 participants with gastric cancer; of these, 12 reported a history of eating chili peppers (exposed). **Controls (disease free)** were 479 participants with no gastric cancer; of these, 88 reported a history of eating chili peppers.

a. (1 point)  
Create a 2 x 2 table from this case control study for analysis.

b. (2 points)  
What is the relative odds (odds ratio) of a history of chili pepper consumption for persons with gastric cancer (cases), compared to persons free of disease? *Show your work.*

c. (2 points)  
In 1-2 sentences, interpret the odds ratio you obtained in “b”, in the context of this study for someone who does not know what it is.

d. (2 points)  
What is the relative odds (odds ratio) of gastric cancer for chili pepper eaters (exposed), compared to non chili pepper eaters? *Show your work.*

e. (3 points)  
The National Institutes of Health (NIH) Surveillance and Epidemiology End Results Program (SEER) estimates that there are 7.4 cases of gastric cancer per 100,000 men and women per year. Given this information, in this study, is it possible to estimate the relative risk (RR) of gastric cancer for chili pepper eaters (exposed), compared to non chili pepper eaters? If yes, what is the relative risk? *Show your work.*