UNIVERSITY OF HOUSTON
PRACTICE MIDTERM EXAMINATION

Term: Fall **Year:** 2009

Student: First Name _____ **Last Name** _____

UH Student ID Number _____

Course Abbreviation and Number	Math 4397/6397
Course Title	Biostatistics
Section(s)	001
Sections of Combined Course(s)	
Section Numbers of Combined Course(s)	
Instructor(s)	B. G. Bodmann

Date of Exam	October, 2009
Time Period	Start time: 2:30pm End time: 3:50pm
Duration of Exam	1 hour 20 mins
Number of Exam Pages (including this cover sheet)	6 pages
Exam Type	Closed book
Additional Materials Allowed	Approved sheet

Notes:

- Please use only the basic mathematical functions on your calculator.
- Show your work on all questions. Simple "yes" or "no" answers will be graded as if blank.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

1. Short answer questions - 9 points each

   a. Is there anything wrong with this statement?

   > You study peas from your recent harvest. The probability that a pea is yellow is 5%. The probability that a pea is wrinkled is 1%. Therefore, there is a 6% probability that a pea is yellow or wrinkled.

   Explain your answer briefly.

   b. Suppose that $X_1, \ldots, X_{10}$ are iid $N(\mu_1, \sigma_1)$ mutually independent of $Y_1, \ldots, Y_{15}$, which are iid $N(\mu_2, \sigma_2)$. What is the variance of $Z = 5(\bar{X} - \bar{Y})$?

   c. When calculating the sample variance, why is it customary to divide the sum of squared deviations from the mean by $(n-1)$ rather than n? (Be brief.)

2. Researchers are interested in a new blood test for diagnosing Kryptonite poisoning (a rare disease). For this test a *positive* result is supposed to indicate the *presence* of Kryptonite poisoning. A study found that 85% of patients who are known to have the disease were *positive* on the blood test. In contrast, 21% of individuals who are known not to have the disease, were *positive* on the blood test.

   a. (9 points) A person has a positive blood test result. Compute an appropriate likelihood ratio and make a statement about the odds of this person having the disease. Do not include any unknown such as prevalence in your statement.

   b. (9 points) Given that the prevalence of the disease is 10% in a subject's population with a positive test result, calculate the probability that the person has the disease.

3. (9 points) Suppose that the US intelligence quotients (IQs) are normally distributed with mean 100 and standard deviation 16. What is the probability that a randomly selected person from this population has an IQ above 132? (Explain your calculation.)

4. (9 points) (Refer to the previous question.) Suppose that in a sample of 5 children 4 have IQs above 132. What is the probability of such an occurrence? What assumptions did you use to calculate this probability?

5. (18 points) Let $X_1, \ldots, X_{10}$ be iid exponentially distributed random variables with mean 1 and $Y_1, \ldots, Y_{10}$ be iid exponential random variables with mean 2. A friend claims that $\bar{X}/\bar{Y}$ should be approximately normal with mean $1/2$ and variance $1/5$. Briefly describe how you could verify this claim using simulation with appropriate commands in R.

The next two questions (6 and 7) involve the following setting: A random sample of 40 United States men contained 25% smokers. Let p be the true prevalence of smoking amongst males in the United States.

6. (9 points) Calculate and interpret a 95% confidence interval for the prevalence of smoking amongst males in the United States.

7. (9 points) Refer to Problem 6. Write down the likelihood function for p (do not draw its graph, just give the expression). Is $p = .35$ or $p = .15$ better supported given the data (why, and by how much)? What value of p is best supported (just give the number, do not derive)?