**Midterm Exam**

**BMI 713**

November 16, 2016

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

**Question 1 [10 points]:**

Given that *X* ~ N(mean=10, var=3), compute the probability that |*X*-10|>7.

**Question 2 [25 points]:**

A pharmaceutical company has developed a drug to lower systolic blood pressure. They are currently in process of testing the effects of the drug, with the hopes of presenting statistical evidence to the FDA that this drug lowers the blood pressure of patients. Therefore, the company has run the following experiment in mice: they injected the drug into 20 mice and measure the mice’s systolic blood pressure after two weeks. The sample average is 115 mmHg and the average of an untreated mouse population is known to be 120 mmHg. Suppose that the standard deviation in this population is known to be 15.

**2a**. [8 points] State the null and alternative hypotheses.

**2b.** [8 points] What type of test is appropriate to test statistical significance? What is the p-value resulting from this test?

**2c**. [9 points] What conclusions can be drawn from this p-value at a significance threshold of 0.05? Should this drug be put on the market? What factors about the experimental design would you consider in making this decision? (No need to do additional calculations)

**Question 3 [20 points]:**

Suppose you strongly suspect there is a relationship between two variables X and Y.  You construct an experimental system capable of setting X to any value and then measuring the associated Y value.  You measure n=9 (X, Y) pairs from your system and decide to use linear regression with model Y = a + bX to look for a relationship.

**3a**. [10 points] Your model produces an estimate of b = -0.1217 and the usual test for b = 0 produces a p-value 0.152.  In one or two sentences, interpret this p-value in terms of the potential relationship between X and Y.

**3b**. [10 points] Your colleagues interpret the result in part (a) as meaning no relationship could possibly exist between X and Y.  Make an argument supporting or refuting this conclusion based on the model used and the data collected.

**Question 4 [20 points]:**

**4a. True or False** [4 points]: Lack of statistical significance means the effect size is small.

**4b. True or False** [4 points]**:** The 95% confidence interval for the sample mean is approximately equal to the sample mean plus and minus two standard deviations

**4c. True or False:** [4 points] The 95% confidence interval is less likely to contain the population mean than the 99% confidence interval.

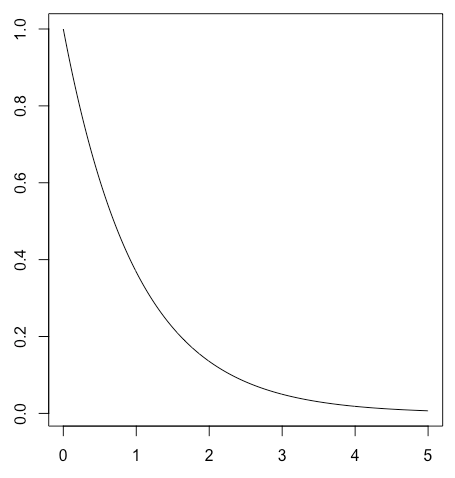
**4d. True or False:** [4 points] The larger the sample size, the more the sampling distribution of sample means will resemble a normal distribution, regardless of the shape of the population distribution

**4e. True or False:** [4 points] The paired t-test assumes that the data in each group are normally distributed.

**Question 5 [25 points]:**

An introductory statistics textbook often states that data must be normally distributed for you to apply the t-test. You would like to investigate whether this is true. To do this, you will perform the t-test on samples drawn from the normal and a non-normal distribution and compare the results.

**5a.** [5 points] Repeat the following test, say, 1000 times: draw 10 samples from the normal distribution N(1,1) and perform the t-test (two-sided) against the null hypothesis that the mean is 1. What fraction of p-values is less than 0.05? What did you expect?

**5b.** [5 points] Repeat the above, except that you draw 10 samples from the gamma distribution using ‘rgamma(10,shape=1,scale=1)’ and perform the t-test against the null hypothesis that the mean is 1. (Note that the gamma distribution with shape=1 and scale=1 is non-normal as shown on the right but has the mean=1). What fraction of p-values is less than 0.05? What did you expect?

**5c.** [5 points] What do you conclude from the results of (a) and (b)?

**5d.** [5 points] Repeat the above, except that you draw 100 samples at each iteration this time. What do you observe?

**5e.** [4 points] Can you explain the different results between n=10 and n=100?

**5f.** [1 point] If you were to give a rule of thumb for the sample size needed for non-normal data, what would it be? Justify your answer.