Name:

Quantitative Understanding in Biology I
2016 Midterm Exam

November 3rd, 2016

Instructions

After the exam, these pages will be separated by question for grading purposes. To ensure that your complete response to each question is considered when grading, please be sure to do the following:

- Hand in all pages that you were given, in order! Even the blank ones at the end, and even if you don't use them (this way we'll know that you didn't use them, and that we didn't lose them).
- Write your name at the top of each side of each page (the exams will be scanned before grading, so we need each page to have your name on it).
- If you do use any of the extra pages, please only respond to one question on a given page, and indicate which question you're addressing at the top of that page.

Question 1

You are rotating in a lab that is working on a cancer therapy, using a mouse model. Your PI has a candidate drug that is looking very promising for inducing apoptosis in cancer cells, but there are concerns that the compound may have a significant side effect impairing cognition.

The lab has decided to do a preliminary check of this hypothesis by comparing the average time that control and treated mice take to learn a maze. From previous work, untreated mice are expected to learn the maze in an average of 8 days, with a standard deviation of $\pm$ 2 days. In lab meeting, a post-doc in the lab has suggested that, in order to have a 50% chance of detecting a delay in learning of 1 day, the lab will need the use of 32 cages in the mouse house (as the animal facility is fondly called in your lab). He shows the following calculation:

$$\text{precision} = 1 \text{ day}$$

$$n = 8\left(\frac{\text{std. deviation}}{\text{precision}}\right)^2 = 8\left(\frac{2 \text{ days}}{1 \text{ day}}\right)^2 = 32$$

Your PI points out that there are only 16 cages available in the mouse house for this experiment, and suggests that the experiment be done with 16 mice, and, if the p-value turns out to be greater than 0.05, the lab will add another 16 mice to the experiment.

Critically evaluate the lab's reasoning and plan. How would you suggest the lab proceed? State what calculations should be done to best test the hypothesis that treatment delays a mouse's ability to learn a maze.

Question 2

In a population of 10,000, 1% of the people have a particular genetic trait that you are interested in. To identify potential recruits for your study, you have two tests available to help screen for the trait.

For Test I:

- A person who has the trait has a 99% chance of testing positive.
- A person who does not have the trait has a 20% chance of testing positive.
- The test costs $100 to administer.

For Test II:

- A person who has the trait has an 80% chance of testing positive.
- A person who does not have the trait has a 1% chance of testing positive.
- The test costs $50 to administer.

Hint: throughout this question, feel free to round intermediate results to make your numerical computations easier. But be sure to show all your work!

What is the *approximate* probability that someone who actually has the trait will be missed if only Test II is used?

What is the *approximate* probability that someone who has a positive result for Test I actually has the trait of interest?

Your research team has decided to adopt a 'two-test' strategy where Test I will be administered first, and, for those people who test positive, Test II will be administered to confirm the positive result. Only those people who have a positive result for both tests will be invited to participate in the study.

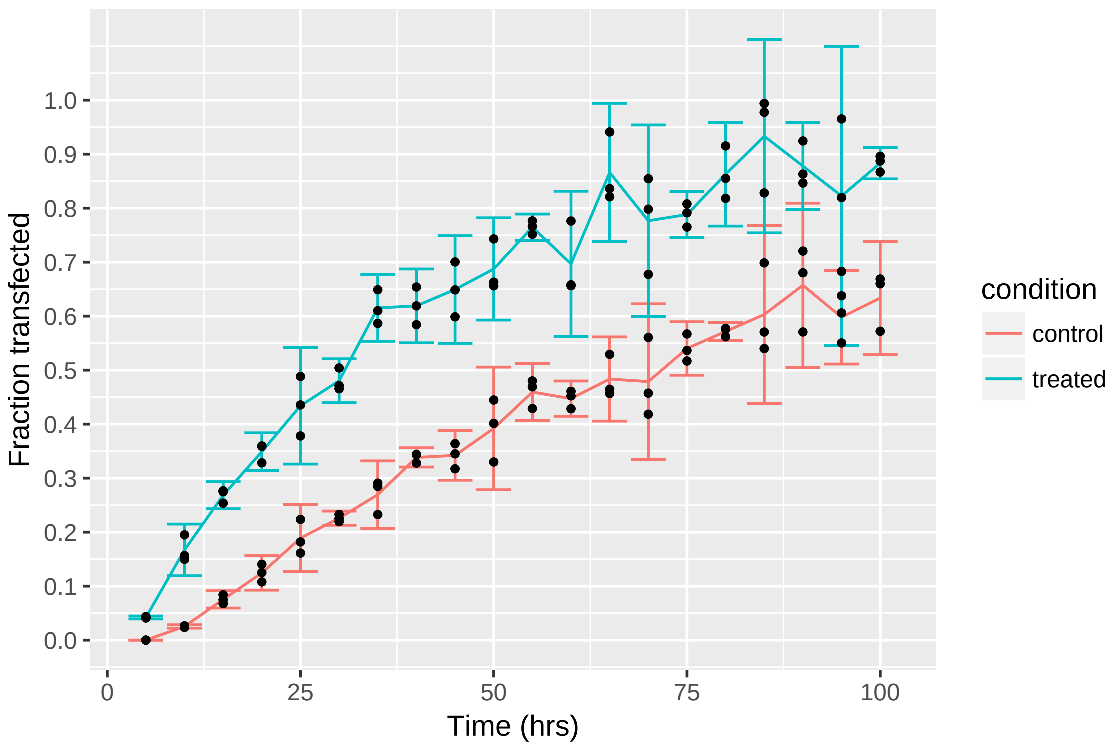
What is the *approximate* probability that someone who has positive results for Test I and then Test II actually has the trait of interest?

What is the *approximate* probability that someone who actually has the trait of interest will not be invited to participate in the study when using the two-test strategy?

What is your estimate of the total cost of screening the population using the two-test strategy?

Question 3

A salesman from the ACME Optical Instrument Company has asked if your lab is interested in a new system to rapidly quantify the time course of transfection of HeLa cells. Using a computer controlled stage and a new AI-driven machine vision system, the method counts the total number of transfected cell nuclei in a field of view, as well as the total number of nuclei, and records the fraction of cells that have been successfully transfected. The salesman says that to get good statistics, the system scans three fields of view for each sample, and uses the observed distribution to compute error bars for the fraction of transfected cells. The system scans a sample every five hours, and a sample graph from their marketing materials looks like the following:



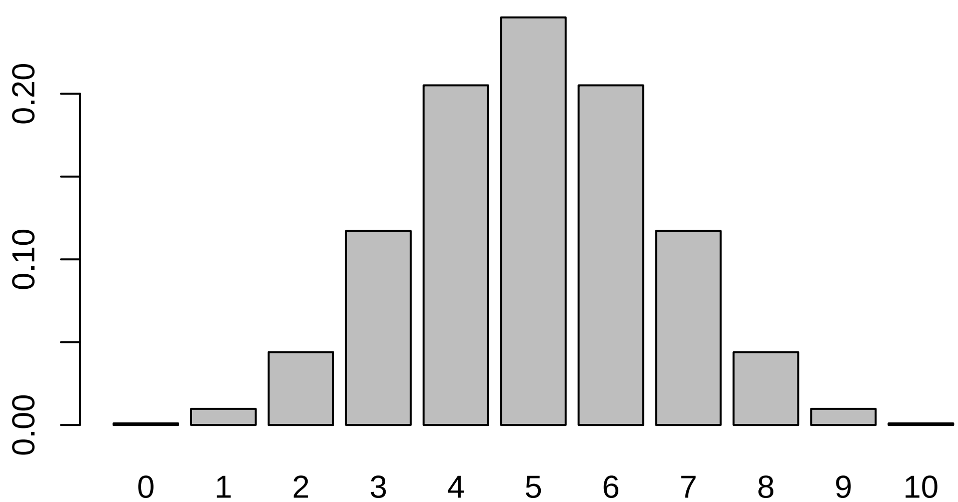
Since taking qBio, you care deeply about what error bars represent. When asked about this, the salesman, after consulting with his technical experts, reports that the error bars' limits are mean $\pm$ $1.96 \cdot \text{sd}$ for the three measurements at each condition and timepoint.

Critically evaluate this analysis. What, if any, changes would you suggest to the plotting and/or approach?

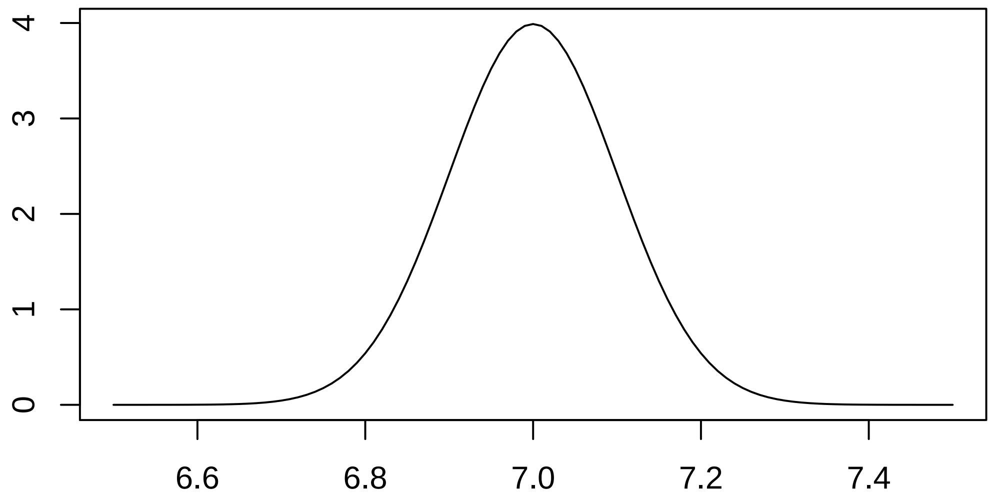
Question 4

Identify what is being represented in the following three plots, and explain your reasoning.

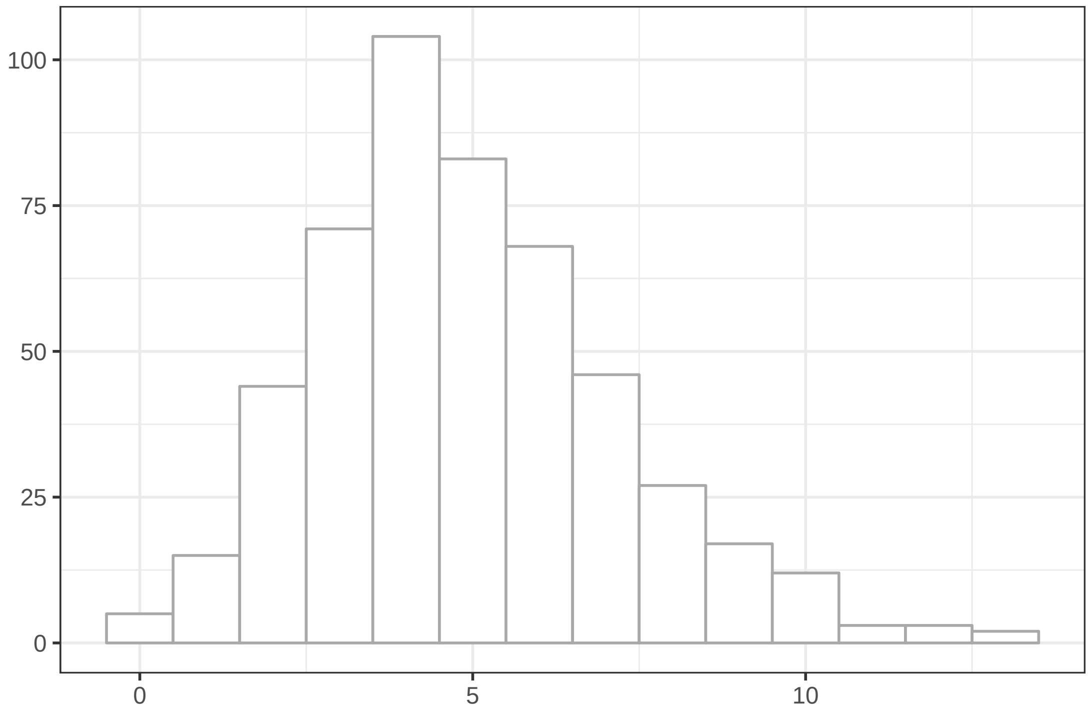
For example, for the figure below you might respond: "This looks like a plot of the PDF for the binomial distribution where $n = 10$ *and* $p = 0.5$*"*



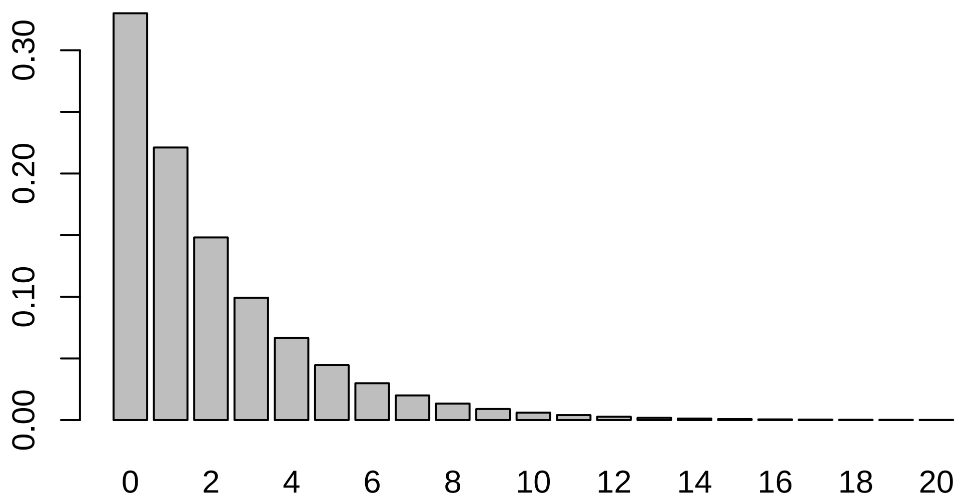
Plot 1



Plot 2



Plot 3



[extra space if you need it]

[extra space if you need it]

[extra space if you need it]

[extra space if you need it]

[extra space if you need it]