BIOS: 4120 Lab 7

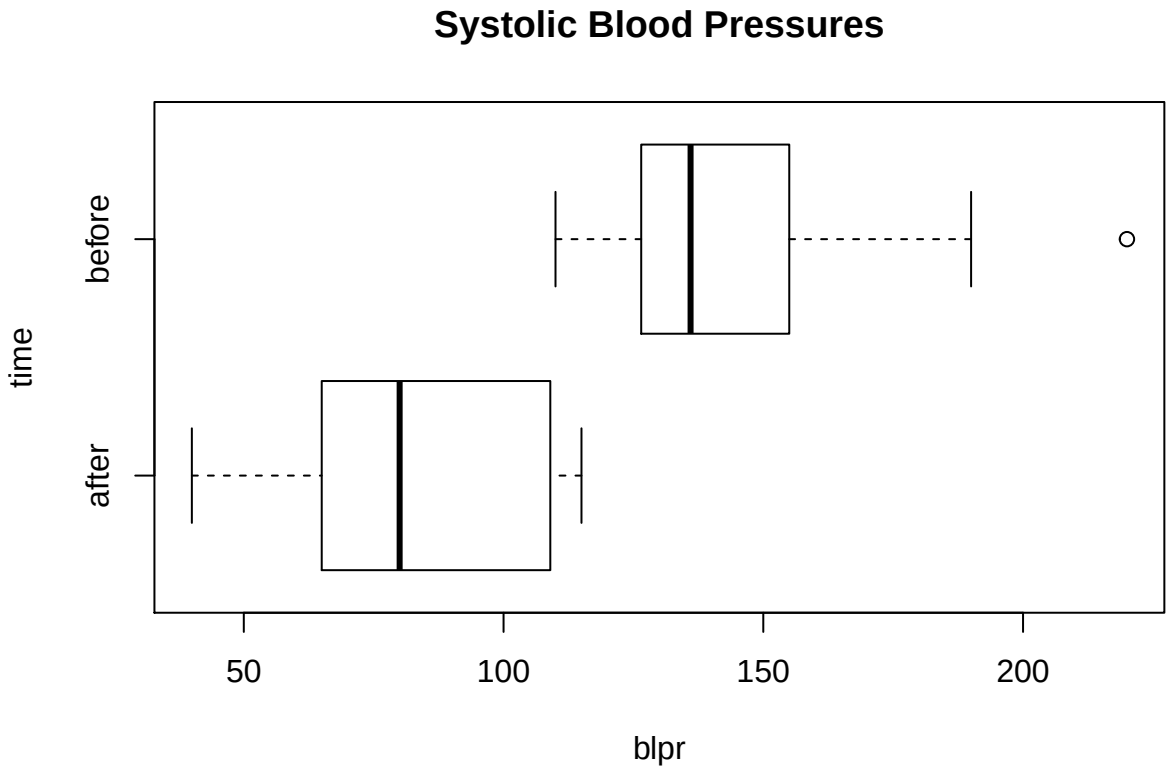
March 3-4, 2020

In today's lab we will review for the quiz, and discuss the binomial distribution and its functions in R.

Quiz Review

Problem 1

Below is a sample of 15 patients' systolic blood pressures prior to having surgery and 15 patients' systolic blood pressures post to having surgery:



a) Estimate the median of blood pressures both before and after surgery.
b) Are there any outliers? If so, identify them and state how they impact the mean and the median?
c) What percentage of individuals had a systolic blood pressure lower than 80 post surgery?
d) 75 percent of individuals had a pre-surgery systolic blood pressure lower than what?

Problem 2 - Probability Review

Suppose the probability that a potato is a Yukon Gold is 1/3.
Suppose the probability that a potato is mashed, given that it was Yukon Gold, is 3/4.
Suppose the probability that a potato is mashed, given that it was not Yukon Gold, is 1/2.

a) What is the probability that a potato is mashed?

b) What is the probability that a potato is both Yukon Gold AND mashed?

c) What is the probability that a potato is Yukon Gold, given that it is mashed?

d) What is the probability that a potato is Yukon Gold OR mashed?

e) Assuming picking potatoes involves independent events, what is the probability that I pick two Yukon Golds in a row (with replacement)?

Problem 3

Below is information about a study concerned with the effect that listening to classical music has on students' test scores. Time spent listening to classical music is given in minutes per week and test scores are the percent that students scored on their biostatistics exams.

```
## Mean Time Listening to Classical Music: 54.44
## Mean Test Scores: 80.36
## Std Dev of Time Listening to Classical Music: 7.3
## Std Dev of Test Scores: 9.9
## Correlation: 0.3919
## Regression Line Slope: 0.531768
```

A. If we have a student who listens to 1 standard deviations less classical music than the average student, how many standard deviations below average would we predict their exam score to be?

B. If we have a student who scores 3 standard deviations above average on their exam, how much time do we predict that they spend listening to classical music per week?

C. If a student listens to 5 minutes more classical music per week than average, what is their predicted test score?

Problem 4

The following data are taken from a study investigating the use of a technique called radionuclide ventriculography as a diagnostic test for detecting coronary artery disease. Assume the participants were randomly selected from the population.

	Disease		
Test	Present	Absent	Total
+	302	80	382
-	179	372	551
Total	481	452	933

a) Find the sensitivity of the test.

b) Find the prevalence of the disease.

c) Find the false positive rate.

d) Find the specificity of the test.

e) Find the predictive value positive. $P(D|T^+)$

f) Find the predictive value negative. $P(D^C|T^-)$

Binomial Distribution (Not on Quiz)

From lecture, we know that when there are two possible outcomes (success/failute) in n trials, the number of ways of one event occurring x times is $\frac{n!}{x!(n-x)!}$.
We also know that, given independence, the probability of an intersection of events is $p^x(1-p)^{1-x}$.
Combining these, we get the formula for the binomial distribution:

$$\frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$$

Using the information about Yukon Gold potatoes from the Probability Review section, let's find the probability that if 3 potatoes are picked, 2 are Yukon Gold. We can calculate this probability using the formula in R:

```
n<-3
x<-2
pi <- 1/3
factorial(n)/(factorial(x)*factorial(n-x)) * pi^x * (1-pi)^(n-x)
```

```
## [1] 0.2222222
```

We can also use R's built-in function to answer this question:

```
dbinom(x=2,size=3,prob=1/3)
```

```
## [1] 0.2222222
```

R's built-in functions can also help us answer other questions. For example, let us now consider picking 10 potatoes and getting 5 Yukon Golds. We can find the probability of this event just like we did earlier:

```
dbinom(x=5,size=10,prob=1/3)
```

```
## [1] 0.1365645
```

However, we may also be interested in finding the probability of seeing an event *as extreme or more extreme* than the one we observed. Since the probability of picking a Yukon Gold is 1/3 and we picked a total of 10 potatoes, we would expect to see about 3.33 Yukon Golds. What we observed (5) is 1.67 greater than what we'd expect, so in order to be *as extreme or more extreme*, we are interested in anything greater or equal to 5 or less than or equal to 1.67. Since the data is discrete, this is the same thing as $P(x \leq 1 \cup x \geq 5)$.

We can calculate this using pbinom(), which finds the probability of being less than or equal to a value. If we want to find the probability of being greater or equal to a number, we tell R to calculate 1-pbinom() of one less than what we're interested in.

```
pbinom(1,size=10,prob=1/3) #Less than or equal to 1.67
```

```
## [1] 0.1040492
```

```
1-pbinom(4,size=10,prob=1/3) #Greater than or equal to 5
```

```
## [1] 0.2131281
```

```
# Total of the Extremes:
pbinom(1,size=10,prob=1/3) + (1-pbinom(4,size=10,prob=1/3))
```

```
## [1] 0.3171773
```

```
# Equivalently:
binom.test(x=5,n=10,p=1/3)$p.value
```

```
## [1] 0.3171773
```

Bonus Practice Problems

For which of the following scenarios could we apply a binomial distribution: Recall: The binomial distribution has the following characteristics:

1) There are a specific number of trials (n), each with a binary outcome
2) The n trials are independent
3) The probability of success (p) is constant with each trial

a) The number of jackpots in 1,000 pulls of a slot machine
b) The number of people who get sick in a 5-person household
c) The number of free throws Lebron James makes in 10 attempts
d) The number of questions a student answers correctly on a multiple choice test (they are not randomly guessing).

Below is data about a fictitious HIV rapid-diagnostic test. Please fill in the rest of the table and answer the following questions.

	Test Result		
Disease	Positive	Negative	Total
Present	2970	30	3000
Absent	11000	539000	550000
Total	13970	539030	553000

a) Find specificity and sensitivity.

b) Assume the prevalence in this population is 0.1%. What is the predictive value positive of this test? Is this a good test? Why or why not?

Will the following data have a higher mean or median? Does that make the data right-skewed or left-skewed?