BIOS: 4120 Lab 7

February 27-28, 2017

In today's lab we will revisit probability, discuss the binomial distribution and its functions in R, and review for quiz 2.

Note: The numbers in this lab are made up rather than based on real data.

Probability Review:

Let event A be that a potato is a Yukon Gold potato.
Suppose the probability that a potato is a Yukon Gold is 1/3.
Let event B be that a potato is used to make mashed potatoes. 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 Yukon Gold, given that it is mashed?
B. What is the probability that a potato is mashed?
C. What is the probability that a potato is both Yukon Gold AND 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?

Binomial Distribution

From lecture, we know that when there are two possible outcomes that occur/don't occur n times, the number of ways of one event occurring k times is $\frac{n!}{k!(n-k)!}$.
We also know that, given independence, the probability of an intersection of events is $p^k(1-p)^{1-k}$.
Combining these, we get the formula for the binomial distribution:

$$\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

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
k<-2
p<-1/3
factorial(n)/(factorial(k)*factorial(n-k)) * p^k * (1-p)^(n-k)
```

```
## [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
```

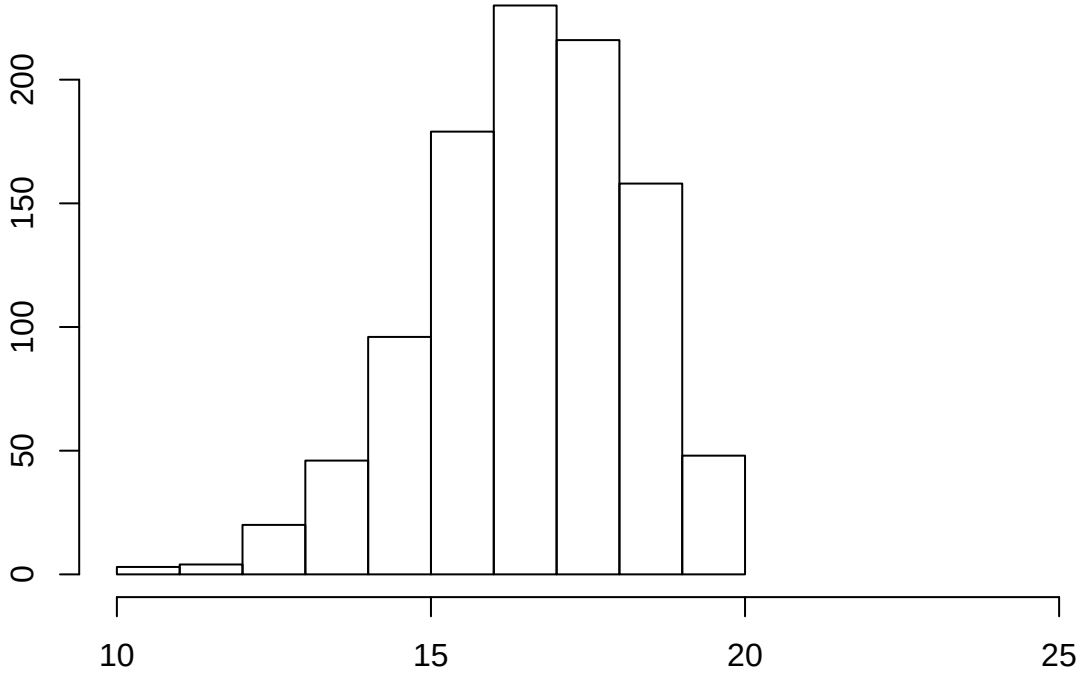
Quiz Review

Problem 1:

```
## Mean of X: 11.33
## Mean of y: 15.88
## Std Dev of X: 2.19
## Std Dev of Y: 2.07
## Correlation: 0.8727
```

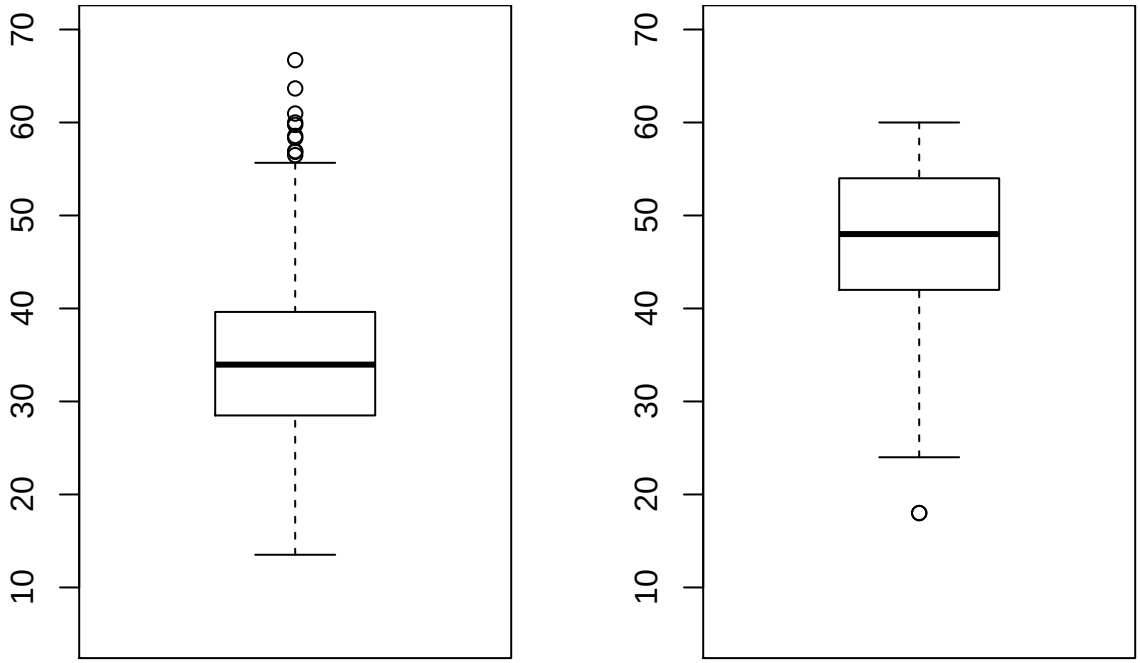
A. If we have an X-value of 14.33, what would we predict the Y-value to be?
B. If we have a Y-value of 11.73, what would we expect the X-value to be?

Problem 2:



What can you say about the distribution (center, shape, spread)? Based on the histogram, what's the approximate standard deviation?

Problem 3:



Compare the two plots (center, shape, spread, outliers). What are the 25th percentiles? The 75th percentiles?

Problem 4:

Suppose 1000 people take a medical screening test. 270 people get a positive test result, 1/3 of which actually have the disease. The prevalence of disease is 0.1. Construct a table with the given information. What are the sensitivity and specificity?