In R, the command to run a binomial distribution is rbinom(). Go to the R documentation and look for the arguments that are required. Then, create a sample of size 1000 using n=8 and p=0.2 as parameters. Assign the created sample to the vector my_binomial. Based on your plot, answer the following questions.

Question 9

(1/1 point)
Write down the simplest code that will allow you to plot a histogram of the sample that you have created. Remember that the name should be **my_binomial**

hist(my_binomial)

Answer: hist(my_binomial)

EXPLANATION

The code in R to run histograms is **hist.** The syntax you need to use is just **hist(x)**, where x is the vector that you want to plot. In this case, the name of that vector is **my_binomial.** Thus, the simplest code is **hist(my_binomial).**

- Exit Survey

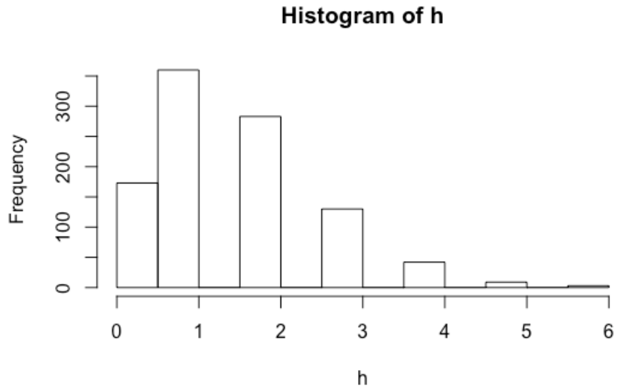
You have used 1 of 2 submissions

Question 10

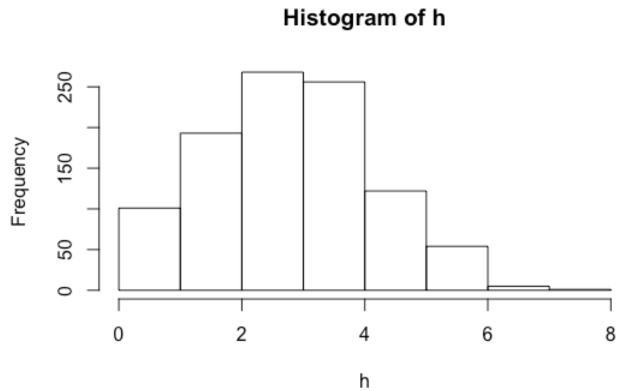
(1/1 point)

Which of the following histograms is closest to the plot that you created?

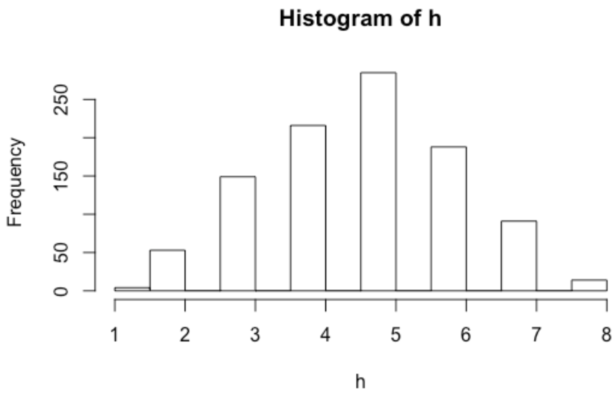
a.



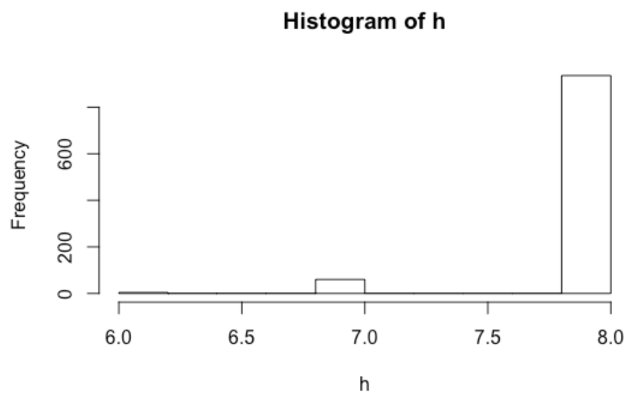
b.



c.



d.



EXPLANATION

The plot in (a) was created with a random sample in which $p = 0.2$. The other plots were created with the following parameters $p = 0.4, 0.6, 0.8, 0.99.$ Thus, since 1000 observations is a number large enough you should obtain a similar histogram to the one in (a).

You have used 1 of 2 submissions

Question 11

(1/1 point)
Now assume that instead of creating the sample with the parameter p equal to 0.2, we set the parameter equal to 1.

What would be the sample mean of this new sample that you have created? (Try to think about the question without coding directly in R)

8

Answer: 8 **or** 1000

And what would be the standard deviation?

0

Answer: 0

0

EXPLANATION

In this case, we know that every trial will be a success (since p=1). Thus, among 8 trials (n=8), we will have 8 successes across the 1000 observations. This implies that the mean of the vector should be 8 and that the standard deviation would be 0.

You have used 1 of 2 submissions

Question 12

(1/1 point)

Now, let's think about the CDF of this variable when p=0.5, and how to construct it in R. We have found a code that constructs the CDF for this variable, but is filled with blanks! Can you fill them for us?

```
my_binomial <- rbinom(1000, 8, p=□)

k <- c(0:8)
cdf <- rep(0.0, times = □)
for (i in 0:8){
  j <- i + 1
  cdf[j] <- (sum(my_binomial<=□)/1000)
}

plot(k, cdf)
```

- a. 0.5; 8; j
- b. 0.2; 9; i
- c. 0.5; 9; i ✔

- d. 0.2; 8; i
- e. 0.5; 9; j

EXPLANATION

Since p=0.5, then the first entry should be that number. We should create the cdf as a vector of nine entries, since the potential values of successes in 8 trials are from 0 to 8. Finally, the relevant parameter to construct the cdf also goes from 0 to 8, and we are doing this in the loop using i.

You have used 1 of 2 submissions

Question 13

(1/1 point)

If a variable z follows a uniform distribution between 0 and 1, what is the value of the CDF evaluated at any value x if $0 \leq x \leq 1$?

x

Answer: x **or** X

EXPLANATION

In the uniform case between 0 and 1, we know that:

$$Pr(z \leq x) = \int_0^x dz = z|_0^x = x$$

You have used 1 of 2 submissions

Question 14

(1/1 point)

Now we will consider two independent random variables that follow a binomial distribution. X follows a binomial distribution with parameters n=8, and p=0.5; while Y follows a binomial distribution with parameters n=8 and p=0.2.

Taking this information into account, what is the probability that the random variable X+Y has a value of zero?

Note: Round to the 6th decimal place. For example, if the answer is 0.00132422, you would input 0.001324

0.000655 ✔ **Answer:** 0.000655

0.000655

EXPLANATION

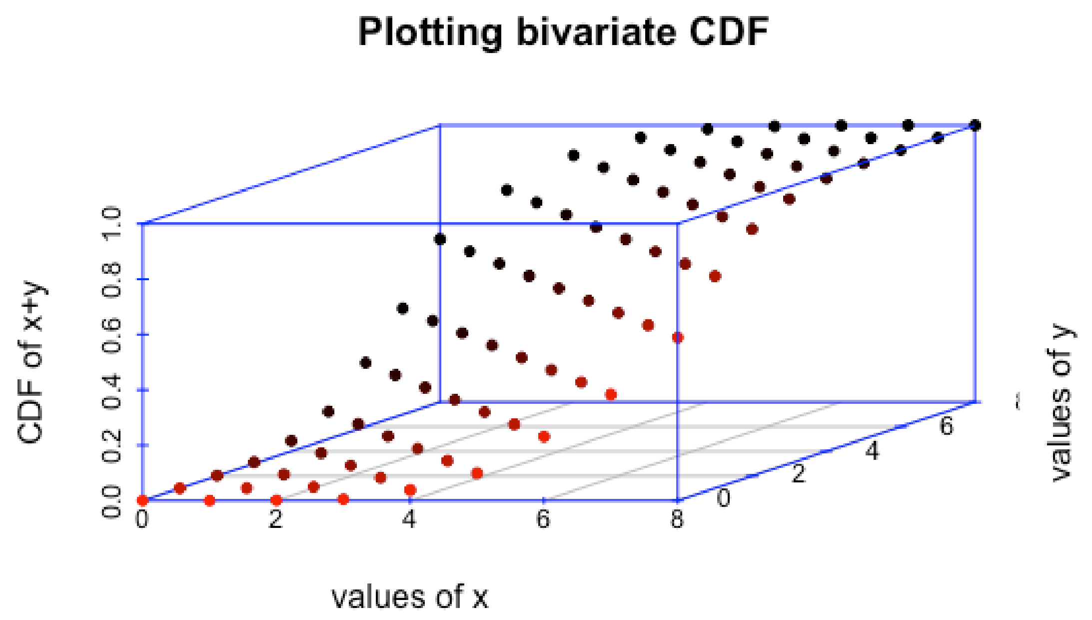
The only way in which X+Y can take a value of zero, is if $X = 0$ and $Y = 0$. Since we know that both variables are independent, then we know that $Pr(X + Y = 0) = Pr(X = 0) * Pr(Y = 0)$. Therefore, we have that 8 choose 0 is equal to one. Then we know that $Pr(X=0)=0.5^8$ and that $Pr(Y=0)=0.8^8$, and from this we have: $Pr(X=0)*Pr(Y=0)=0.00065536 \approx 0.000655$

You have used 2 of 2 submissions

One of the cool things about R is that users have developed different packages that you can use. In this example, we are going to use one of this packages called scatterplot3d. We want to plot in a three-dimensional plane the cumulative distribution of X+Y. Take a look at the code and try to understand it. Furthermore, give it a try on your computer.

```
install.packages("scatterplot3d")
library("scatterplot3d")

x <- rbinom(1000, 8, 0.5)
y <- rbinom(1000, 8, 0.5)

k <- c(rep(0:8, times=9), rep(0:8, each=9))
k <- matrix(k, ncol=2, byrow=FALSE)
z <- k[,1]+k[,2]

cdf <- rep(0.0, times = 81)

for (i in 1:81){
  cdf[i] <- sum(x+y <= z[i])/1000
}

scatterplot3d(k[, 1], k[, 2], cdf, highlight.3d = TRUE, col.axis
= "blue",
              main="Plotting bivariate CDF", pch = 20,
ylab="values of y", xlab="values of x",
              zlab = "CDF of x+y")
```

The code produces the following figure:



Question 15

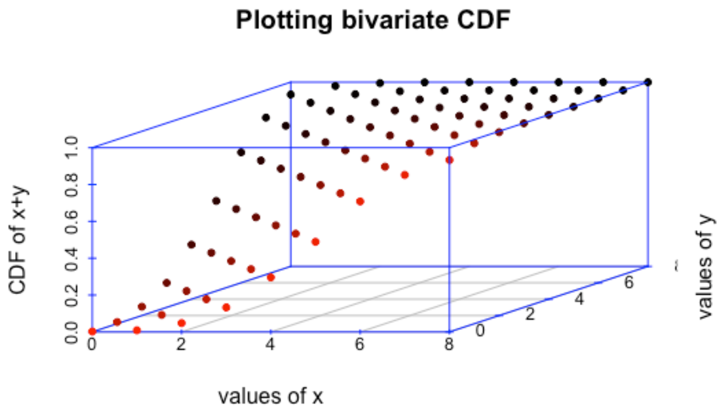
(1/1 point)

There is a mistake in the above code compared to what the original problem states. In particular, while the the code indicates Y as binomially distributed with n=8, it incorrectly sets p=0.5. Instead, you want a plot with p=0.2.

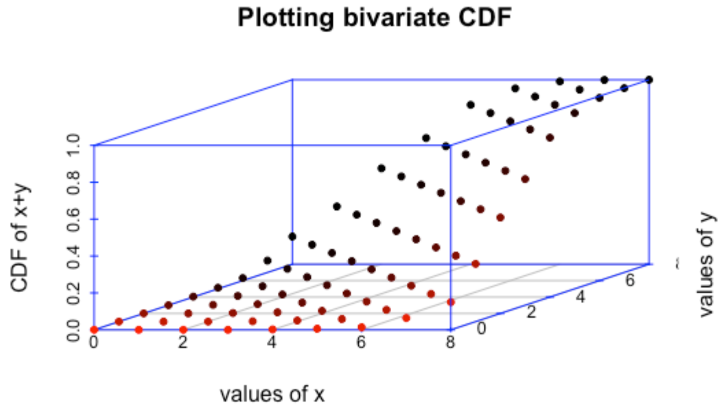
One of your friends has given you two additional plots - one is correct with p=0.2 and the other with p=0.8. Your friend, however, forgets which plot is which. Take a look at the two plots below and determine which one corresponds to **p=0.2.**

You can try to solve the question analytically or use the help of R to check your answer.

a.



b.



EXPLANATION

Compared to p=0.8, when p=0.2, we have that lower values of X+Y are more likely. Thus, the CDF of X+Y should always be higher for p=0.2 than when p=0.8. This concept is called first order stochastic dominance. In particular we are going to have that for all values of k, $Pr(X+Y \leq k|p=0.8) \leq Pr(X+Y \leq k|p=0.2)$.

You have used 1 of 1 submissions