lab4_estes

Andrew Estes

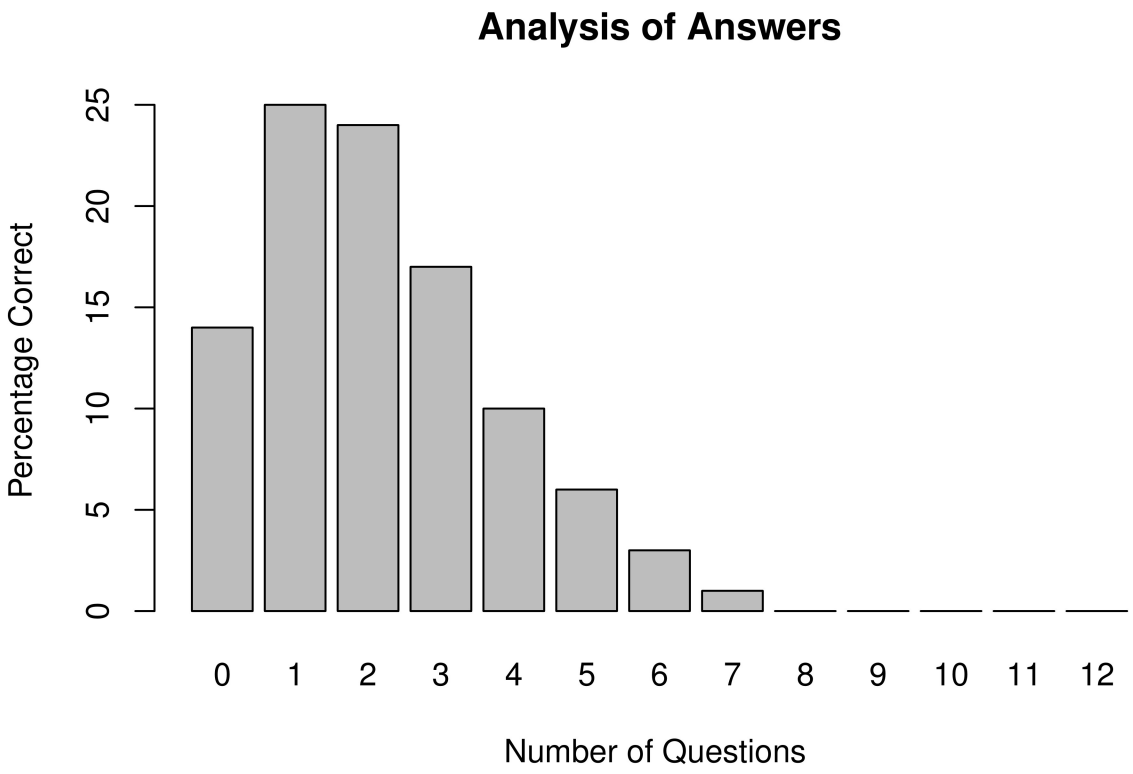
9/18/2021

1

In R, create a numeric vector containing these percentages, and then use the barplot command to create a graph of the distribution of "number correct." Make sure the graph axes are labeled correctly, and add an explanatory title.

```
number_correct <- c(.14, .25, .24, .17, .1, .06, .03, .01, 0, 0, 0, 0, 0)
answers <- c(0:12)
number_correct <- number_correct * 100

barplot(number_correct ~ answers,
        horiz = FALSE,
        main = "Analysis of Answers",
        ylab = "Percentage Correct",
        xlab = "Number of Questions ")
```



2

Describe the shape of the graph you obtained

The graph above is heavily skewed to the right, meaning correct answers are hard to come by. It is so hard in fact that nearly 75% of all correct answers only answered 1/3rd of the answers correctly.

3A

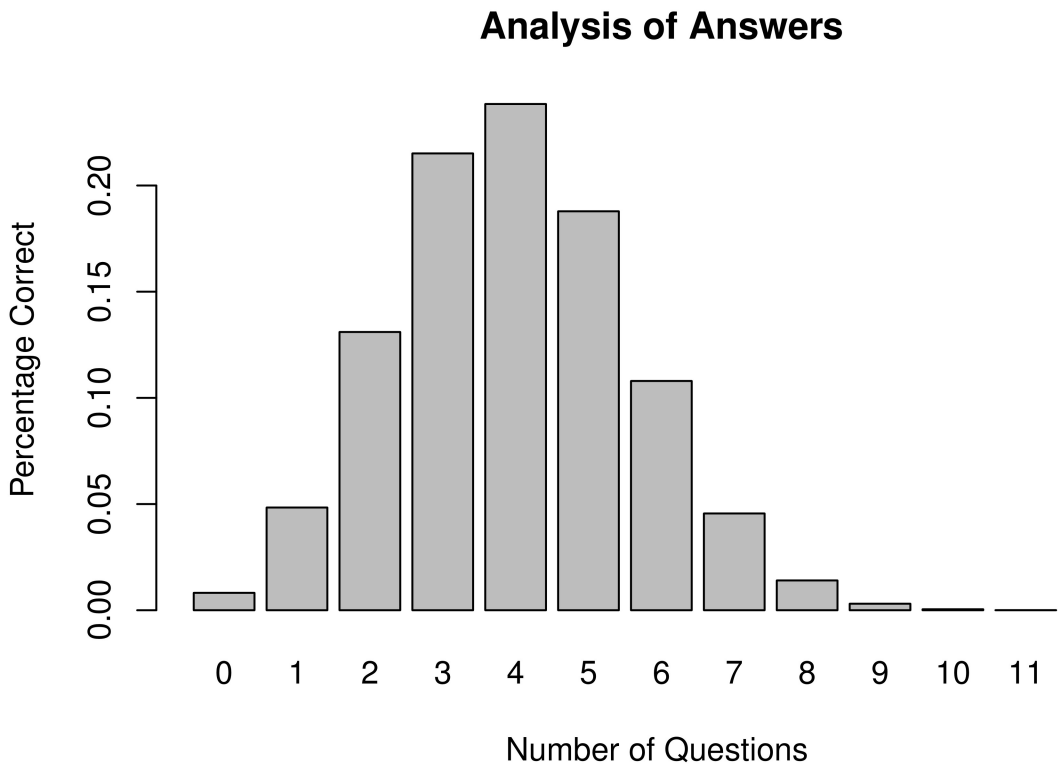
There are 12 questions, each with three options, and the chimpanzees really do choose randomly, The number of correct responses X should be described by a binomial probability distribution. Indicate the values of n and p for this binomial distribution. P = .33 because there is a 1 in 3 chance the Chimp selects the correct answer N = 12

3B

```
x <- seq(0, 12, by = 1)
y <- dbinom(x, 12, 0.33)

barplot(y ~ x,
        horiz = FALSE,
        main = "Analysis of Answers",
        ylab = "Percentage Correct",
        xlab = "Number of Questions ")
```

Use the dbinom function to create a new vector of probabilities reflecting the chimp's expected performance on the quiz. As in Question 1, plot the probability that a chimp will get every-



where from 0 to 12 correct.

3C

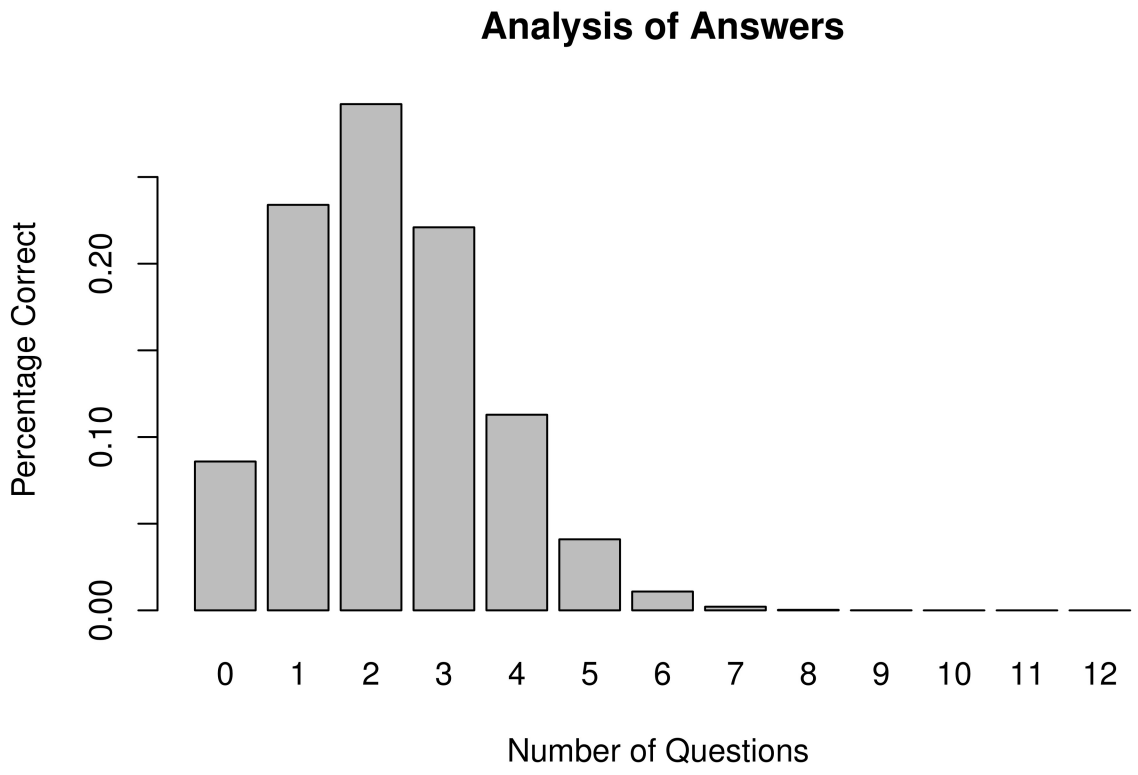
Describe how humans compare to chimps by comparing distribution center, spread and shape in your answer. Humans skewed right to the extreme.Chimps have a slight right skew. Humans somewhat followed the normal distribution/bell curve. Chimps followed the normal distribution/.bell curve Humans center at 2. Chimps center at 4 (.33 * 12).

4A

```
x2 <- seq(0, 12, by = 1)
y2 <- dbinom(x, 12, 0.185)

barplot(y2 ~ x2,
        horiz = FALSE,
        main = "Analysis of Answers",
        ylab = "Percentage Correct",
        xlab = "Number of Questions ")
```

Use the dbinom command again to create a barplot of the expected distribution of correct human answers, if human responses were also described by a binomial distribution.



4B

```
a <- pbinom(12, 12, .185)
b <- pbinom(5, 12, .185)
c <- a - b
c
```

Write code that uses the pbinom() function to calculate P(X > 4) for the binomial description of the number of correct human responses.

```
## [1] 0.01329958
```

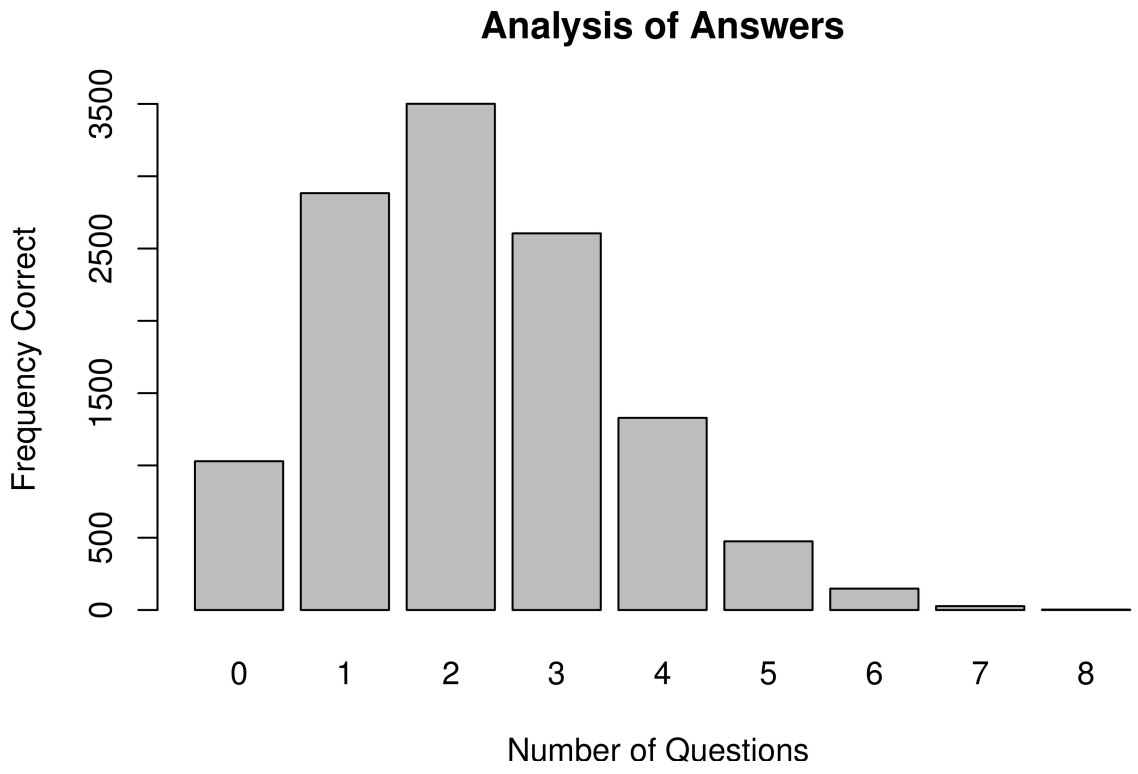
4C

How closely does your graph from Question 4a resemble the actual distribution of number of select answers from Question 1? In other words, for which number of correct answers were the actual human probabilities higher or lower than the binomial distribution? They match up surprisingly well. The DBINOM resulted in higher scores.

5A

```
five <- rbinom(12000, 12, .185)
a <- table(five)
barplot(a,
        horiz = FALSE,
        main = "Analysis of Answers",
        ylab = "Frequency Correct",
        xlab = "Number of Questions "
)
```

Gapminder states that they surveyed 12,000 people. Write code to create a simulated set of results using the rbinom command. [Remember that p = 0.185.] You should get a vector whose length is 12,000, with each entry containing the simulated number of correct answers from one respondent. After generating this simulated data, graph it using a combination of barplot and table commands to obtain a relative frequency distribution.[Note: You could use the hist command with this raw data, but it's formatting is a bit finicky.]



5B

What I would like you to do is to push the "green triangle" button in R Studio to run your code chunk from Question 5a many times. You should get a different random sample every time,

provided you haven't included a set.seed() command. Does the graph of your random sample ever look like the actual human results (i.e., bars with heights that match the actual human results)? In other words, does it seem likely that the real human results could have been one random results from a binomial distribution with the appropriate success probability? Yes, it could have been a random result within the parameters of a binomial distribution with a probability success of around 0.185.

6

Finally, let's think about what's going on here. The binomial distribution makes several assumptions in order to calculate the probability of a certain number of successes: • There are two possible outcomes: success or failure. (In this case correct or incorrect answers.) • There are a fixed number of trials. (In this case, a trial is one person answering one question, and our random variable is the number of correct that a person got out of that total number of trials.) • The probability of success is constant over all trials. • Results of one trial are independent of other trials. (In other words, knowing the success/failure result for one trial or some trials won't help you predict results of other trials.) One explanation for any differences between the real human results and your simulated results could lie in the fact that one or more of these assumptions is not actually satisfied. Thinking about this realworld situation, briefly explain why or why not these assumptions might be satisfied. Your explanation should be sure to explicitly connect the general assumption to the specifics of this real-world example.

There is one major assumption not satisfied: there is not a fixed number of trials. Anybody can take the test as often as they would like. This means people may increase their score by re-taking the test multiple times which indicates the actual rate of correct answers is less than 0.185. There are other issues with the test not included in these assumptions such as demographics. What if the majority of human test-takers are middle school students? Surely they score differently than college students? What about socio-economic background - is there a correlation between socio-economic status and answers correct? There are many other questions and variables that are missing that can better define the results. However, given what we have - even with the error in at least one of the four major assumptions - the binomial distribution relatively followed the pattern of actual data so I must begrudgingly conclude the assumptions are correct.