\*\*For this homework, I would like you to think about and play around with probability distributions. It is necessary to understand these distributions and how they relate to statistical significance (i.e., passing a threshold that allows you to reject the null hypothesis).\*\*

(1) Your friend is convinced that he has extra-sensory perception (ESP), and you are convinced that he does not. So you design a test. You shuffle a full deck of cards (i.e., 52 cards) ; you ask him to say what the top card is; you then turn the card over to see if he is correct. After each guess, you put the card back in the deck and repeat the procedure. You do this a bunch of times, and he only gets a couple right. So you say, "See. You don't have ESP. If you were psychic, you would have been able to tell me the correct card each time." To which he responds, "Well, I didn't say it was perfect. But I still have ESP. After all, I got a few right." So you play again. This time you play the game 200 times, and you carefully record the result each time. (You and your friend have far too much time on your hands!) How many times would your friend have to get the correct answer in order for you to reject the null hypothesis that he does not have (helpful) ESP? (Hint: You will first have to calculate the probability of success on a single trial. Obviously, this is not .5, as it is when you flip a coin.)

Probability of success is determined by dividing the number of trials by tokens. Since there are fifty-two cards in a deck, there are fifty-two tokens.

1/52=0.01923077

dbinom(1,200,0.01923077)

[1] 0.08069122

qbinom(.05,200,0.01923077,lower.tail = FALSE)

[1] 7

qbinom(.05/2,200,0.01923077,lower.tail = FALSE)

[1] 8

pbinom(8,200,0.01923077,lower.tail = FALSE)

[1] 0.01611731

2\*pbinom(8,200,0.01923077,lower.tail = FALSE)

[1] 0.03223461

2\*pbinom(9,200,0.01923077,lower.tail = FALSE)

[1] 0.01151091

With 0.03223461 or 3% probability of the likelihood on the upper and lower tail from eight and more cards, this allows us to reject the null hypothesis that he doesn’t have psychic powers and, rather, assume that he just might have ESP. However, just looking at the spectrum of eight cards or more, the probability of this kid knowing the cards is 2% or 0.01611731 by calculating pbinom without multiplying it by two.

The two times of the pbinom allows us to look at one end rather than two ends—a one directional hypothesis.

(2) Your other friend sees you playing this game and says that he wants to try next because he has an especially weird kind of ESP. It is more like a little voice in his head that tells him the answers to things. (You need to get some new friends!) Some days, his little voice is helpful, and tells him the correct answers. But other days, the voice messes with him and only gives him the wrong answers. Could you use your test to investigate this? How many times would your friend have to get the correct answer in order for you to reject the null hypothesis that he does not have this weird ESP? (Remember, you don't know whether his voice is being helpful or not on this particular day.)

qbinom(.05,200,0.01923077,lower.tail = FALSE)

[1] 7

qbinom(.05/2,200,0.01923077,lower.tail = FALSE)

[1] 8

> qbinom(.05/2,200,0.01923077,lower.tail = TRUE)

[1] 1

> qbinom(.05,200,0.01923077,lower.tail = TRUE)

[1] 1

This problem requires us to show a non-directional result, both higher and lower ends. Just as we said earlier that a person would have to get eight or more cards right for us to suspect that they just might have ESP, we can be quite assured that getting one card right only assures us that this kid doesn’t have any kind of ESP or “voice” giving him the answers. (However, the kid might want to get checked out by a psychologist anyways if he is hearing voices that are not really there.)

BONUS:

Evaluate the observed frequency of the word "senator" in the Brown corpus based on the probability of this word occurring in the CELEX database [Brown corpus frequency = 40 (corpus size = 1000000); CELEX database frequency = 267 (corpus size = 18580121)].

40/1000000

[1] 4e-05

.00004

(i) What is the probability of obtaining a value at the observed frequency of 40 or higher in the Brown corpus?

qbinom(.05,1000000,.00004,lower.tail = FALSE)

[1] 51

pbinom(51,1000000,.00004,lower.tail = FALSE )

[1] 0.03873714

2\*pbinom(51,1000000,.00004,lower.tail = FALSE )

[1] 0.07747427

dbinom(40,1000000,.00004)

[1] 0.0629483

The Brown corpus has the probability of eight or 0.07747427 to obtain a value with a frequency of 40 or more.

(ii) What does this say about the frequency of the word "senator" in the Brown corpus?

267/18580121

[1] 1.43702e-05

dbinom(267,18580121,.0000143702)

[1] 0.02440743

2\*pbinom(51,18580121,.0000143702,lower.tail = FALSE )

[1] 2

> qbinom(.05,18580121, .0000143702,lower.tail = FALSE)

[1] 294

> qbinom(.05/2,18580121, .0000143702,lower.tail = FALSE)

[1] 299

Looking at the dbinom, we can see that “senator” has a higher frequency to occur in the Brown’s corpus. The density probability is six for Brown’s corpus in contrast to two likely to occur in CELEX’s corpus. Though the frequency is 267 for CELEX, that is out of more than 18 million, and less likely to occur in CELEX, in contrast to 40 out of 1 million in Brown’s corpus.