All right we're back with doing some stuff in R. Now we could perform binomial calculations by hand. But it's a little bit more work. and a little bit more error-prone than I really care for us to, to use.

Now in most of my classes I use TI calculator, Texas Instruments calculator usually TI 84 or 83. You may remember some of the TI line back from your college days that if you had classes that required them. They're still pretty much a staple in education these days. And they're convenient to use, but because we're going to have access to software, I'm going to just use R. I will say this that I'll give pretty much an outline for how we use R. Below I have a an output of the help file, but we'll jump over to R in a minute.

So in R, we are going to have different distributions and they are going to have certain names. When we're looking at distributions, they're all going to have a prefix. So there's going to be D, P, Q and R. And I'll show you how to bring up the R Help here in a moment. But this is the output here. So what all is here?

Well, in each help file, you'll see a kind of a description of the function. You'll see some usages, so some modifications on the functions. Because this is for a distribution, we're going to have these four separate functions. All distributions will have that. When we start dealing with the normal and T distributions in the next course. There'll be dnorm, pnorm, qnorm, rnorm. DT, PT, QT and RT; we'll see all of those pop up again. The next input are the arguments of the function. So basically what each thing represents. I'll step through each of those here.

Basically all of the functions will have a first input that differs based on the function. A size, that's the number of trials, that's going to actually be n in our case. The probability trial is going to be P. Now the next input that we're going to worry about is this lower tail. So it's true by default. If it's true, then what we're going to have is a lower tail probability. You can also have an upper tail probability. I'm going to tend to avoid that. We'll talk about it a little bit. But we'll, again, we'll worry about that a little bit later. I'll mention it, but I'm going to tend to rely on this.

So this is basically the CDF indicator. So, so we'll deal with that here in a few moments. And then a little bit further down, you'll see some details. And you can also get some examples.

So let's hop over to R and look at some of these probabilities. So I'll hop over here in just a moment. Alright let's go ahead and hop over. I just need to make sure things were kind of cleaned up a little bit. But we'll go ahead and take a look at things now. Okay so I'm just going to go over here to R. So a few things to note. [Get some of stuff out of the way. There we go.]

So this might look a little different, it's just the setup I use. I changed the color the background color a little bit, here. [Let me fix something here real quick.] [Okay, I'm getting some stuff out of the way. You might see some kind of weird things blocking, but I think I got it taken care up here. I had to pause recording for a moment.] Anyway, so this is going to be the R console right here. [Sorry, something popped up there.]

So this is the R console right here. This is where you can just type in commands for some things. So, like for example, we'll go through this in another video but if, I wanted to help on the binomial function, what I would do is I would first hit a question mark and then type the name of a function. So for example if I want to know something about the binomial distribution I have dbinom.

Now, you'll often get an autocomplete kind of pop up here. So you can tab complete kind of like you can, I guess like most terminals and things like that. You can hit tab. You can do sometimes do a double tab and it'll give you some, some examples. Like if I want to know some about the mean, or median. I can hit tab, and it'll give me pretty much all of the matching names. So if I want to know something about the mean function, I can hit enter and down here, you'll see some, basically the help file, pop up. If you hit this button right here, next to the printer, basically opens a new window. So you can kind of look through the help that I displayed earlier, and they usually give some examples as well.

So one thing I did want to note here real quick, and again, I'll pull this up in another video. But if you do example and then open parentheses and type the name of a function, you will get basically the help file. It will give you an idea of how to do some things. So, for example, if I, if I do this-this is a little bit outdated-this left arrow as an assignment operator. But if I type that in, basically what it will do, is it will give me values from 0 to 10 and then 50. So if I type X, then it gives me 0 to 10 in steps of one and then 50 added on. And then if I do mean of X that will give me the mean. And then there's trimmed mean which we didn't really talk about much so. But anyway that is or maybe we've talked about it briefly. But so that's some stuff in R. I'm going to hit this little broom here. It's going to clear the console.

So this is just a console. There's a terminal which is more of, kind of like, a like a Linux thing. So I can do all my stuff, if I want to see my present working directory, all that other stuff. [Well I guess that present working direction doesn't work but anyway anyway, so that's what I did, that's another project I'm working on, but whatever it is.] So I'll go back over to console. If you want to write a script, which is something you'll see in our labs, up here you'll see a plus sign. And so you can do an R script. There are notebooks kind of similar to like a Jupyter notebook in Python or something like that and then there's some other stuff you can do. But I just wanted R script. And if I open that up, it'll just be basically a window that I can type some code in. I don't really need it right now I'm just going to get rid of this, and we'll stick to the console.

So let me hop back over to our notes and let's start looking at an example. So Tay-Sachs disease is a rare but fatal genetic condition. [unknown phrase] Disease of genetic origin, pardon me, occurring chiefly in infants and children and especially those of Jewish or Eastern European extraction. If there was a couple that are both carriers of the disease or the child of theirs has a probability of 0.25 of being born with the disease. If such a couple has four children what is the PMF?

Okay, so what we really need is, at the end the day, what we want is, we want the PMF. So that's going to be the values of X, and then the probability with which X is observed. So X, we have to figure out what values can it take on, and then we can start filling in these probabilities. So first thing is that this is the binomial distribution. It's going to feel a little weird thinking about that, in that way. But we have a set number of trials ahead of time. So, so number of trials is basically going to be our births. So four births. We're going to assume that the trials are independent. That's an assumption we're going to make just to make things a little bit easier. Are they really independent? I'm not a biologist so I don't know.

But we're going to make that assumption. And then are, so this is basically going to be the value of N. So four is basically going to be the value of N. So N is going to be 4. And then, here this is going to sound a little weird, but a success, is going to be having the disease, which is going to mean that P is going to be 0.25. And that's actually all we need. We just need to know what the value of N is and what the value of P is. Once we know N and P, we can do these calculations. So X can go from 0, there could be 0 children that have the disease. 1, 2, 3, 4, or all four children could have the disease.

So we need to find these probabilities. In order to do that we're going to have to use the binom PDF or the binomial function the PMF of the binomial function. So we're going to use R for this. So I want to find 0, the probability of 0. Now the general notation to calculate these probabilities, we're going to have to use the Dbinom function. So I'm going to scroll up here just a little bit.

We're going to use this guy, right there, okay. So there are basically four inputs we have to look at. The X is going to be the number, right. So at 0, 1, 2,3 or 4 size is my sample size N prob is my probability of success which is going to be P or 0.25. This log function here is basically should we report report the log of our probabilities. We don't really have a need to do that. So because we don't have a need to really deal with that, we're just going to leave it alone. This equals false line here is going to basically be we call a default value. So unless we change it, it's going to assume that log is equal to false.

So we're just going to ignore that last input. But we need that dbinom. So to get this probability, I probably should have written this higher up, I'm going to write it in green just to differentiate it. So what we're going to have to use, is we're going to use the dbinom function. My first input X it's going to be 0. The size input is going to be N which is 4. And then the prob input is key and that's going to be 0.25. Lets hop over to R and see what we get. So I'm going to type in dbinom by itself. And open parentheses. And what you can do is you can actually hit the tab button and it will prompt you for all the inputs.

I like to do this at least starting out with a prob [I'm having to cough a lot, so I'm having to pause the recording.] But anyway, when I'm back here, when I hit tab it's going to give me a prompt for each of the inputs. So if I hit tab again, it'll say X equals. You don't have to do this, as long as you put things in the right order, you have to have the argument, but I think starting out this is a good habit to get into.

So X for us is going to be 0 We're going to hit comma tab. Size is going to be my N, that's going to be 4. Comma tab prob, and that's going to be 0.25. Now I don't have to do the last input. I could if I wanted to be pure about it, but I don't have to. But when I hit Enter, there's my probability. So I'm just going to round to three decimal places to make it a little bit cleaner so 0.316. So over here my probability is 0.316. Okay alright. Okay so so we got the probability for one, for 0, pardon me. Now we're going to look at the probability of one.

So I'm going to hop back over here to R. And I'm going to do this again. I'm going to calculate dbinom. dbinom, but you can arrow down and hit tab and it will auto fill. So X in this case is going to be 1. My size, again, is going to be 4. And prob is going to be 0.25. Let's hit Enter. All right so, so there we go. So 0.42- 0.421875- 0.422 is what I'm looking for. There we go. Okay so 0.422. I'll draw some lines in here to establish demarcation between the different rows. Let's do it for 2. So let's see so dbinom. X is going to be [whoops] 2. N is going to be 4. Prob is going to be 0.25, so 0.211. [excuse me]

Alright let's do the last couple. So here's an interesting feature that you can use an R. If I hit the up arrow key it'll just cycle back through things, up and down I can cycle through. So I go back here and just change the 3 and then we can also do 4. So we'll just kind of finish that out. So so three decimal places so 0.047 for 3, so 0.047 and then 4 was 0.004. And then there we go. We have it all. So let's look at what each one of these represents. So let's take the case of 2 here. So let's look at this right here. So this tells me that the probability of 2 is about 0.211. Again, I'm rounding to 3 decimal places. But that's okay.

So this tells me that if I have a family, or a couple, where both are carriers of this disease, and they have four kids, and the probability that 2 of those 4 children that have that are going to have the disease is 0.211, so well but more in a 21% chance. And this is for any 2 kids. So, it can be the first and second, the first and third, the first and fourth and so on. And so we could do that calculation of how many there are but turns out there are six combinations of that, it's 4 choose 2. But that's not really important, or paramount right now. The important thing is we know how to use the binomial function in the R environment.

So I'm going to go ahead and do example three in this video just because it all fits together fairly nicely. So let's take a look. So we have the case where sickle-cell anemia, again, another disease, where a person has two copies of a certain recessive gene. People with one copy of the gene are called carriers. Carriers don't have disease but can pass that gene on to their children. So if we have a child to parents who are both carriers, a probability of 0.25 again, of having sickle cell anemia. Well let's say we have a medical study of 18 children from families where both parents are carriers. So what's the probability that 4 or more of the children have sickle cell anemia?

Okay, so now we're going to have to look at a new function in R, a different function for the binomial. But before we go too far, we're going to again note that this is the binomial scenario. We have at this time N. [Oops. Back to black here.] N is 18. And again P is 0.25 here. So this is my N, and this is my P. So I can move on with that. So the probability of 4 or more...the probability of 4 or more. Now there's a couple ways we could go about this. I can calculate all the probabilities of 4, 5, 6, 7 all the way up to 18 as we did in the example 2. That's going to take a long time and then I have to add them all up. So that's just going to be a lot of work.

So we want to use a CDF function here. Remember CDF sums from the left. So I have to write this as a complement. So this is going to be equal to 1 minus the complement of this. And this is going to be less than 4. Or we can say 1 minus 3 or less. So there's actually two ways we could do this. So we're still going to have the 1 minus. But how do I do the probability of 3 or less?

Well let's jump back over to R real quick. And I'm going to just clear this, just to alleviate confusion. But let's look at the binom help file again. [retype, dbinom] So let me open that up and let's take a look here. So here dbinom is basically the probability of an individual value. pbinom is what we're going to actually use here. pbinom is basically going to be the cumulative distribution function.

That's what we're going to use. So it's got almost the same inputs. But we're going to want to use the pbinom. And here instead of an X it's a Q for a quantile what we'll talk about that, kind of be patient, a little bit later. But basically it's going to be my 3. It's going to be the same as my X above. And then the size is going to be 4 and the prob, again, is going to be 0.25. Now if we go back over to the Help, Of course I minimized it, But if we go back over here to the Help, [or let me actually do this over in the, I forgot I had it up here further.]

So this lower tail right here that we talked about earlier, this is going to come in to play. It's true by default. And so it's basically going to do this right here. It's kind of hard to see, but the probability that X is less than or equal to some value. We can do a right tail probability by setting it equal to false. And we'll see how we can kind of do that in a different way here. But let's hop back over R. So what we want is we want to do the pbinom function. So pbinom. Actually we want 1 minus that, right. 1 minus pbinom. And we said that Q was going to be 3. The size is going to be a 4. And my probability is 0.25. Now the lower tail is true by default. So I'm just going to leave it and hit Enter. And it's pretty low probability 0.0039 or 0.004, for us. So this is going to be pretty low probability. This is [I'm running out of space here, unfortunately didn't really think ahead] This is going to be about 0.004. So pretty low chance that's going to happen. Because, if you think about it we're looking at 18 children. So the probability that we're going to have 4 or more is going to be pretty low. But we'll kind of justify that a little bit later.

Another way we could have done it, I just want to note here, is that we could have used the pbinom function again and had Q be equal to this time, it would still be 3. The size is going to be 4 again. Prob is 0.25. And if we set lower tail to false, that we have to it either as a capital F, or you have to write it all in capital letters. We get the same probability. So we get the same thing. It's just we don't have to do the 1 minus. So when I say lower tail is equals to false, it's basically doing the 1 minus part of it. I prefer the former rather than the latter.

Okay let's look at part B. What is the probability that fewer than three children have the disease sickle cell anemia. So this is the probability of fewer than three. And so we wanted to write that as an inclusive of a value. So it's going to be the probability of two or less. This is kind of equivalent to saying the probability that X is less than 3 is equal to the probability that X is less than or equal to 2, right. If I have fewer than three I can't have half a child with the disease, right. So we have to go with integer steps And I want to be in the less than or equal to form. Because if we go back over to the Help [ let's scroll up again through real quick.] That's the form that we need. So I need a less than or equal to symbol there. So what we can do here is we can actually just use the pbinom function, and Q, again, is going to 2. The size, is going to be 4 and my prob is going to be 0.25. So we can get an answer here. So I'll hop over here. So pbinom Q is 2, because its fewer than three, So 2 or less. Size is 4 and probability is 0.25. And we get a pretty large value, and I just realized I made a mistake. I'll have to re-record this whole video.