

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu).

**JOHN TSITSIKLIS:** So today we're going to finish with the core material of this class. That is the material that has to do with probability theory in general. And then for the rest of the semester we're going to look at some special types of models, talk about inference. Well, there's also going to be a small module of core material coming later.

But today we're basically finishing chapter four. And what we're going to do is we're going to look at a somewhat familiar concept, the concept of the conditional expectation. But we're going to look at it from a slightly different angle, from a slightly more sophisticated angle. And together with the conditional expectation we will also talk about conditional variances. It's something that we're going to denote this way.

And we're going to see what they are, and there are some subtle concepts that are involved here. And we're going to apply some of the tools we're going to develop to deal with a special type of situation in which we're adding random variables. But we're adding a random number of random variables.

OK, so let's start talking about conditional expectations. I guess you know what they are. Suppose we are in the discrete world.  $x, y$ , or discrete random variables. We defined the conditional expectation of  $x$  given that I told you the value of the random variable  $y$ . And the way we define it is the same way as an ordinary expectation, except that we're using the conditional PMF.

So we're using the probabilities that apply to the new universe where we are told the value of the random variable  $y$ . So this is still a familiar concept so far. If we're dealing with the continuous random variable  $x$  the formula is the same, except that here we have an integral, and we have to use the conditional density function of  $x$ .

Now what I'm going to do, I want to introduce it gently through the example that we talked about last time. So last time we talked about having a stick that has a certain length. And we take that stick, and we break it at some point that we choose uniformly at random. And let's

denote why the place where we chose to break it.

Having chosen  $y$ , then we're left with a piece of the stick. And I'm going to choose a place to break it once more uniformly at random between 0 and  $y$ . So this is the second place at which we are going to break it, and we call that place  $x$ .

OK, so what's the conditional expectation of  $x$  if I tell you the value of  $y$ ? I tell you that capital  $Y$  happens to take a specific numerical value. So this capital  $Y$  is now a specific numerical value,  $x$  is chosen uniformly over this range. So the expected value of  $x$  is going to be half of this range between 0 and  $y$ . So the conditional expectation is little  $y$  over 2.

The important thing to realize here is that this quantity is a number. I told you that the random variable took a certain numerical value, let's say 3.5. And then you tell me given that the random variable took the numerical value 3.5 the expected value of  $x$  is 1.75. So this is an equality between numbers.

On the other hand, before you do the experiment you don't know what  $y$  is going to turn out to be. So this little  $y$  is the numerical value that has been observed when you start doing the experiments and you observe the value of capital  $Y$ . So in some sense this quantity is not known ahead of time, it is random itself. So maybe we can start thinking of it as a random variable.

So to put it differently, before we do the experiment I ask you what's the expected value of  $x$  given  $y$ ? You're going to answer me well I don't know, it depends on what  $y$  is going to turn out to be. So the expected value of  $x$  given  $y$  itself can be viewed as a random variable, because it depends on the random variable capital  $Y$ .

So hidden here there's some kind of statement about random variables instead of numbers. And that statement about random variables, we write it this way. By thinking of the expected value, the conditional expectation, as a random variable instead of a number. It's a random variable when we do not specify a specific number, but we think of it as an abstract object.

The expected value of  $x$  given the random variable  $y$  is the random variable  $y$  over 2 no matter what capital  $Y$  turns out to be. So we turn and take a statement that deals with equality of two numbers, and we make it a statement that's an equality between two random variables. OK so this is clearly a random variable because capital  $Y$  is random. What exactly is this object? I didn't yet define it for you formally. So let's now give the formal definition of this object that's

going to be denoted this way.

The conditional expectation of  $x$  given the random variable  $y$  is a random variable. Which random variable is it? It's the random variable that takes this specific numerical value whenever capital  $Y$  happens to take the specific numerical value little  $y$ . In particular, this is a random variable, which is a function of the random variable capital  $Y$ . In this instance, it's given by a simple formula in terms of capital  $Y$ . In other situations it might be a more complicated formula.

So again, to summarize, it's a random. The conditional expectation can be thought of as a random variable instead of something that's just a number. So in any specific context when you're given the value of capital  $Y$  the conditional expectation becomes a number. This is the realized value of this random variable. But before the experiment starts, before you know what capital  $Y$  is going to be, all that you can say is that the conditional expectation is going to be  $1/2$  of whatever capital  $Y$  turns out to be.

This is a pretty subtle concept, it's an abstraction, but it's a useful abstraction. And we're going to see today how to use it. All right, I have made the point that the conditional expectation, the random variable that takes these numerical values is a random variable. If it is a random variable this means that it has an expectation of its own. So let's start thinking what the expectation of the conditional expectation is going to turn out to be.

OK, so the conditional expectation is a random variable, and in general it's some function of the random variable  $y$  that we are observing. In terms of numerical values if capital  $Y$  happens to take a specific numerical value then the conditional expectation also takes a specific numerical value, and we use the same function to evaluate it. The difference here is that this is an equality of random variables, this is an equality between numbers.

Now if we want to calculate the expected value of the conditional expectation we're basically talking about the expected value of a function of a random variable. And we know how to calculate expected values of a function. If we are in the discrete case, for example, this would be a sum over all  $y$ 's of the function whose expected value we're taking times the probability that  $y$  takes on a specific numerical value.

OK, but let's remember what  $g$  is. So  $g$  is the numerical value of the conditional expectation of  $x$  with  $y$ . And now when you see this expression you recognize it. This is the expression that we get in the total expectation theorem. Did I miss something? Yes, in the total expectation

theorem to find the expected value of  $x$ , we divide the world into different scenarios depending on what  $y$  happens. We calculate the expectation in each one of the possible worlds, and we take the weighted average.

So this is a formula that you have seen before, and you recognize that this is the expected value of  $x$ . So this is a longer, more detailed derivation of what I had written up here, but the important thing to keep in mind is the moral of the story, the punchline. The expected value of the conditional expectation is the expectation itself.

So this is just our total expectation theorem, but written in more abstract notation. And it comes handy to have this more abstract notation, as as we're going to see in a while. OK, we can apply this to our stick example. If we want to find the expected value of  $x$  how much of the stick is left at the end?

We can calculate it using this law of iterated expectations. It's the expected value of the conditional expectation. We know that the conditional expectation is  $y$  over 2. So expected value of  $y$  is  $l$  over 2, because  $y$  is uniform so we get  $l$  over 4. So this gives us the same answer that we derived last time in a rather long way.

All right, now that we have mastered conditional expectations, let's raise the bar a little more and talk about conditional variances. So the conditional expectation is the mean value, or the expected value, in a conditional universe where you're told the value of  $y$ . In that same conditional universe you can talk about the conditional distribution of  $x$ , which has a mean-- the conditional expectation-- but the conditional distribution of  $x$  also has a variance.

So we can talk about the variance of  $x$  in that conditional universe. The conditional variance as a number is the natural thing. It's the variance of  $x$ , except that all the calculations are done in the conditional universe. In the conditional universe the expected value of  $x$  is the conditional expectation. This is the distance from the mean in the conditional universe squared. And we take the average value of the squared distance, but calculate it again using the probabilities that apply in the conditional universe.

This is an equality between numbers. I tell you the value of  $y$ , once you know that value for  $y$  you can go ahead and plot the conditional distribution of  $x$ . And for that conditional distribution you can calculate the number which is the variance of  $x$  in that conditional universe. So now let's repeat the mental gymnastics from the previous slide, and abstract things, and define a random variable-- the conditional variance. And it's going to be a random variable because we

leave the numerical value of capital  $Y$  unspecified.

So ahead of time we don't know what capital  $Y$  is going to be, and because of that we don't know ahead of time what the conditional variance is going to be. So before the experiment starts if I ask you what's the conditional variance of  $x$ ? You're going to tell me well I don't know, it depends on what  $y$  is going to turn out to be. It's going to be something that depends on  $y$ . So it's a random variable, which is a function of  $y$ .

So more precisely, the conditional variance when written in this notation just with capital letters, is a random variable. It's a random variable whose value is completely determined once you learned the value of capital  $Y$ . And it takes a specific numerical value. If capital  $Y$  happens to get a realization that's a specific number, then the variance also becomes a specific number. And it's just a conditional variance of  $y$  over  $x$  in that universe.

All right, OK, so let's continue what we did in the previous slide. We had the law of iterated expectations. That told us that expected value of a conditional expectation is the unconditional expectation. Is there a similar rule that might apply in this context?

So you might guess that the variance of  $x$  could be found by taking the expected value of the conditional variance. It turns out that this is not true. There is a formula for the variance in terms of conditional quantities. But the formula is a little more complicated. It involves two terms instead of one. So we're going to go quickly through the derivation of this formula. And then, through examples we'll try to get some interpretation of what the different terms here correspond to.

All right, so let's try to prove this formula. And the proof is sort of a useful exercise to make sure you understand all the symbols that are involved in here. So the proof is not difficult, it's 4 and 1/2 lines of algebra, of just writing down formulas. But the challenge is to make sure that at each point you understand what each one of the objects is.

So we go into formula for the variance affects. We know in general that the variance of  $x$  has this nice expression that we often use to calculate it. The expected value of the squared of the random variable minus the mean squared. This formula, for the variances, of course it should apply to conditional universes. I mean it's a general formula about variances. If we put ourselves in a conditional universe where the random variable  $y$  is given to us the same math should work.

So we should have a similar formula for the conditional variances. It's just the same formula, but applied to the conditional universe. The variance of  $x$  in the conditional universe is the expected value of  $x$  squared-- in the conditional universe-- minus the mean of  $x$ -- in the conditional universe-- squared. So this formula looks fine.

Now let's take expected values of both sides. Remember the conditional variance is a random variable, because its value depends on whatever realization we get for capital  $Y$ . So we can take expectations here. We get the expected value of the variance.

Then we have the expected value of a conditional expectation. Here we use the fact that we discussed before. The expected value of a conditional expectation is the same as the unconditional expectation. So this term becomes this. And finally, here we just have some weird looking random variable, and we take the expected value of it.

All right, now we need to do something about this term. Let's use the same rule up here to write down this variance. So variance of an expectation, that's kind of strange, but you remember that the conditional expectation is random, because  $y$  is random. So this thing is a random variable, so this thing has a variance.

What is the variance of this thing? It's the expected value of the thing squared minus the square of the expected value of the thing. Now what's the expected value of that thing? By the law of iterated expectations, once more, the expected value of this thing is the unconditional expectation. And that's why here I put the unconditional expectation.

So I'm using again this general rule about how to calculate variances, and I'm applying it to calculate the variance of the conditional expectation. And now you notice that if you add these two expressions  $c$  and  $d$  we get this plus that, which is this. It's equal to-- these two terms cancel, we're left with this minus that, which is the variance of  $x$ . And that's the end of the proof.

This one of those proofs that do not convey any intuition. This, as I said, it's a useful proof to go through just to make sure you understand the symbols. It starts to get pretty confusing, and a little bit on the abstract side. So it's good to understand what's going on.

Now there is intuition behind this formula, some of which is better left for later in the class when we talk about inference. The idea is that the conditional expectation you can interpret it as an estimate of the random variable that you are trying to-- an estimate of  $x$  based on

measurements of  $y$ , you can think of these variances as having something to do with an estimation error. And once you start thinking in those terms an interpretation will come about. But again as I said this is better left for when we start talking about inference.

Nevertheless, we're going to get some intuition about all these formulas by considering a baby example where we're going to apply the law of iterated expectations, and the law of total variance. So the baby example is that we do this beautiful experiment of giving a quiz to a class consisting of many sections. And we're interested in two random variables. So we have a number of students, and they're all allocated to sections. The experiment is that I pick a student at random, and I look at two random variables.

One is the quiz score of the randomly selected student, and the other random variable is the section number of the student that I have selected. We're given some statistics about the two sections. Section one has 10 students, section two has 20 students. The quiz average in section one was 90. Quiz average in section two was 60. What's the expected value of  $x$ ? What's the expected quiz score if I pick a student at random?

Well, each student has the same probability of being selected. I'm making that assumption out of the 30 students. I need to add the quiz scores of all of the students. So I need to add the quiz scores in section one, which is 90 times 10. I need to add the quiz scores in that section, which is 60 times 20. And we find that the overall average was 70. So this is the usual unconditional expectation.

Let's look at the conditional expectation, and let's look at the elementary version where we're talking about numerical values. If I tell you that the randomly selected student was in section one what's the expected value of the quiz score of that student? Well, given this information, we're picking a random student uniformly from that section in which the average was 90. The expected value of the score of that student is going to be 90.

So given the specific value of  $y$ , the specific section, the conditional expectation or the expected value of the quiz score is a specific number, the number 90. Similarly for the second section the expected value is 60, that's the average score in the second section. This is the elementary version. What about the abstract version? In the abstract version the conditional expectation is a random variable because it depends. In which section is the student that I picked?

And with probability  $1/3$ , I'm going to pick a student in the first section, in which case the

conditional expectation will be 90, and with probability  $2/3$  I'm going to pick a student in the second section. And in that case the conditional expectation will take the value of 60. So this illustrates the idea that the conditional expectation is a random variable. Depending on what  $y$  is going to be, the conditional expectation is going to be one or the other value with certain probabilities.

Now that we have the distribution of the conditional expectation we can calculate the expected value of it. And the expected value of such a random variable is  $1/3$  times 90, plus  $2/3$  times 60, and it comes out to equal 70. Which miraculously is the same number that we got up there.

So this tells you that you can calculate the overall average in a large class by taking the averages in each one of the sections and weighing each one of the sections according to the number of students that it has. So this section had 90 students but only  $1/3$  of the students, so it gets a weight of  $1/3$ .

So the law of iterated expectations, once more, is nothing too complicated. It's just that you can calculate overall class average by looking at the section averages and combine them. Now since the conditional expectation is a random variable, of course it has a variance of its own. So let's calculate the variance. How do we calculate variances? We look at all the possible numerical values of this random variable, which are 90 and 60. We look at the difference of those possible numerical values from the mean of this random variable, and the mean of that random variable, we found that's its 70. And then we weight the different possible numerical values according to their probabilities.

So with probability  $1/3$  the conditional expectation is 90, which is 20 away from the mean. And we get this squared distance. With probability  $2/3$  the conditional expectation is 60, which is 10 away from the mean, has this squared distance and gets weighed by  $2/3$ , which is the probability of 60. So you do the numbers, and you get the value for the variance equal to 200.

All right, so now we want to move towards using that more complicated formula involving the conditional variances. OK, suppose someone goes and calculates the variance of the quiz scores inside each one of the sections. So someone gives us these two pieces of information. In section one we take the differences from the mean in that section, and let's say that the various turns out to be a number equal to 10 similarly in the second section. So these are the variances of the quiz scores inside individual sections.

The variance in one conditional universe, the variance in the other conditional universe. So if I pick a student in section one and I don't tell you anything more about the student, what's the variance of the random score of that student? The variance is 10.

I know why, but I don't know the student. So the score is still a random variable in that universe. It has a variance, and that's the variance. Similarly, in the other universe, the variance of the quiz scores is this number, 20.

Once more, this is an equality between numbers. I have fixed the specific value of  $y$ . So I put myself in a specific universe, I can calculate the variance in that specific universe.

If I don't specify a numerical value for capital  $Y$ , and say I don't know what  $Y$  is going to be, it's going to be random. Then what kind of section variance I'm going to get itself will be random. With probability  $1/3$ , I pick a student in the first section in which case the conditional variance given what I have picked is going to be 10. Or with probability  $2/3$  I pick  $y$  equal to 2, and I place myself in that universe. And in that universe the conditional variance is 20.

So you see again from here that the conditional variance is a random variable that takes different values with certain probabilities. And which value it takes depends on the realization of the random variable capital  $Y$ . So this happens if capital  $Y$  is one, this happens if capital  $Y$  is equal to 2.

Once you have something of this form-- a random variable that takes values with certain probabilities-- then you can certainly calculate the expected value of that random variable. Don't get intimidated by the fact that this random variable, it's something that's described by a string of eight symbols, or seven, instead of just a single letter. Think of this whole string of symbols there as just being a random variable. You could call it  $z$  for example, use one letter.

So  $z$  is a random variable that takes these two values with these corresponding probabilities. So we can talk about the expected value of  $Z$ , which is going to be  $1/3$  times 10,  $2/3$  times 20, and we get a certain number from here.

And now we have all the pieces to calculate the overall variance of  $x$ . The formula from the previous slide tells us this. Do we have all the pieces? The expected value of the variance, we just calculated it. The variance of the expected value, this was the last calculation in the previous slide. We did get a number for it, it was 200. You add the two, you find the total variance.

Now the useful piece of this exercise is to try to interpret these two numbers, and see what they mean. The variance of  $x$  given  $y$  for a specific  $y$  is the variance inside section one. This is the variance inside section two. The expected value is some kind of average of the variances inside individual sections.

So this term tells us something about the variability of this course, how widely spread they are within individual sections. So we have three sections, and this course happens to be-- OK, let's say the sections are really different. So here you have undergraduates and here you have post-doctoral students. And these are the quiz scores, that's section one, section two, section three. Here's the mean of the first section. And the variance has something to do with the spread. The variance in the second section has something to do with the spread, similarly with the third spread.

And the expected value of the conditional variances is some weighted average of the three variances that we get from individual sections. So variability within sections definitely contributes something to the overall variability of this course. But if you ask me about the variability over the entire class there's a second effect. That has to do with the fact that different sections are very different from each other. That these courses here are far away from those scores.

And this term is the one that does the job. This one looks at the expected values inside each section, and these expected values are this, this, and that. And asks a question how widely spread are they? It asks how different from each other are the means inside individual sections? And in this picture it would be a large number because the difference section means are quite different.

So the story that this formula is telling us is that the overall variability of the quiz scores consists of two factors that can be quantified and added. One factor is how much variability is there inside individual sections? And the other factor is how different are the sections from each other? Both effects contribute to the overall variability of this course.

Let's continue with just one more numerical example. Just to get the hang of doing these kinds of calculations, and apply this formula to do a divide and conquer calculation of the variance of a random variable. Just for variety now we're going to take a continuous random variable.

Somebody gives you a PDF of this form, and they ask you for the variance. And you say oh

that's too complicated, I don't want to do integrals. Can I divide and conquer?

And you say OK, let me do the following trick. Let me define a random variable,  $y$ . Which takes the value 1 if  $x$  falls in here, and takes the value 2 if  $x$  falls in the second interval. And let me try to work in the conditional world where things might be easier, and then add things up to get the overall variance.

So I have defined  $y$  this particular way. In this example  $y$  becomes a function of  $x$ .  $y$  is completely determined by  $x$ . And I'm going to calculate the overall variance by trying to calculate all of the terms that are involved here. So let's start calculating.

First observation is that this event has probability  $1/3$ , and this event has probability  $2/3$ . The expected value of  $x$  given that we are in this universe is  $1/2$ , because we have a uniform distribution from 0 to 1. Here we have a uniform distribution from 1 to 2, so the conditional expectation of  $x$  in that universe is  $3/2$ .

How about conditional variances? In the world where  $y$  is equal to 1  $x$  has a uniform distribution on a unit interval. What's the variance of  $x$ ? By now you've probably seen that formula, it's  $1/12$ .  $1/12$  is the variance of a uniform distribution over a unit interval.

When  $y$  is equal to 2 the variance is again  $1/12$ . Because in this instance again  $x$  has a uniform distribution over an interval of unit length. What's the overall expected value of  $x$ ? The way you find the overall expected value is to consider the different numerical values of the conditional expectation. And weigh them according to their probabilities.

So with probability  $1/3$  the conditional expectation is  $1/2$ . And with probability  $2/3$  the conditional expectation is  $3/2$ . And this turns out to be  $7/6$ .

So this is the advance work we need to do, now let's calculate a few things here. What's the variance of the expected value of  $x$  given  $y$ ? Expected value of  $x$  given  $y$  is a random variable that takes these two values with these probabilities.

So to find the variance we consider the probability that the expected value takes the numerical value of  $1/2$  minus the mean of the conditional expectation. What's the mean of the conditional expectation? It's the unconditional expectation. So it's  $7/6$ . We just did that calculation. So I'm putting here that number,  $7/6$  squared. And then there's a second term with probability  $2/3$ , the conditional expectation takes this value of  $3/2$ , which is so much away from the mean, and we get this contribution.

So this way we have calculated the variance of the conditional expectation, this is this term. What is this? Any guesses what this number is? It's  $1/12$ , why? The conditional variance just happened in this example to be  $1/12$  no matter what. So the conditional variance is a deterministic random variable that takes a constant value. So the expected value of this random variable is just  $1/12$ . So we got the two pieces that we need, and so we do have the overall variance of the random variable  $x$ .

So this was just an academic example in order to get the hang of how to manipulate various quantities. Now let's use what we have learned and the tools that we have to do something a little more interesting. OK, so by now you're all in love with probabilities. So over the weekend you're going to bookstores to buy probability books.

So you're going to visit a random number bookstores, and at each one of the bookstores you're going to spend a random amount of money. So let  $n$  be the number of stores that you are visiting. So  $n$  is an integer-- non-negative random variable-- and perhaps you know the distribution of that random variable.

Each time that you walk into a store your mind is clear from whatever you did before, and you just buy a random number of books that has nothing to do with how many books you bought earlier on the day. It has nothing to do with how many stores you are visiting, and so on. So each time you enter as a brand new person, and buy a random number of books, and spend a random amount of money.

So what I'm saying, more precisely, is that I'm making the following assumptions. That for each store  $i$ -- if you end up visiting the  $i$ -th store-- the amount of money that you spend is a random variable that has a certain distribution. That distribution is the same for each store, and the  $x_i$ 's from store to store are independent from each other. And furthermore, the  $x_i$ 's are all independent of  $n$ . So how much I'm spending at the store-- once I get in-- has nothing to do with how many stores I'm visiting.

So this is the setting that we're going to look at.  $y$  is the total amount of money that you did spend. It's the sum of how much you spent in the stores, but the index goes up to capital  $N$ . And what's the twist here? It's that we're dealing with the sum of independent random variables except that how many random variables we have is not given to us ahead of time, but it is chosen at random.

So it's a sum of a random number of random variables. We would like to calculate some quantities that have to do with  $y$ , in particular the expected value of  $y$ , or the variance of  $y$ . How do we go about it?

OK, we know something about the linearity of expectations. That expectation of a sum is the sum of the expectations. But we have used that rule only in the case where it's the sum of a fixed number of random variables. So expected value of  $x$  plus  $y$  plus  $z$  is expectation of  $x$ , plus expectation of  $y$ , plus expectation of  $z$ . We know this for a fixed number of random variables. We don't know it, or how it would work for the case of a random number.

Well, if we know something about the case for fixed random variables let's transport ourselves to a conditional universe where the number of random variables we're summing is fixed. So let's try to break the problem divide and conquer by conditioning on the different possible values of the number of bookstores that we're visiting. So let's work in the conditional universe, find the conditional expectation in this universe, and then use our law of iterated expectations to see what happens more generally.

If I told you that I visited exactly little  $n$  stores, where little  $n$  now is a number, let's say 10. Then the amount of money you're spending is  $x_1$  plus  $x_2$  all the way up to  $x_{10}$  given that we visited 10 stores. So what I have done here is that I've replaced the capital  $N$  with little  $n$ , and I can do this because I'm now in the conditional universe where I know that capital  $N$  is little  $n$ . Now little  $n$  is fixed.

We have assumed that  $n$  is independent from the  $x_i$ 's. So in this universe of a fixed  $n$  this information here doesn't tell me anything new about the values of the  $x$ 's. If you're conditioning random variables that are independent from the random variables you are interested in, the conditioning has no effect, and so it can be dropped.

So in this conditional universe where you visit exactly 10 stores the expected amount of money you're spending is the expectation of the amount of money spent in 10 stores, which is the sum of the expected amount of money in each store. Each one of these is the same number, because the random variables have identical distributions. So it's  $n$  times the expected value of money you spent in a typical store.

This is almost obvious without doing it formally. If I'm telling you that you're visiting 10 stores, what you expect to spend is 10 times the amount you expect to spend in each store individually. Now let's take this equality here and rewrite it in our abstract notation, in terms of

random variables. This is an equality between numbers. Expected value of  $y$  given that you visit 10 stores is 10 times this particular number.

Let's translate it into random variables. In random variable notation, the expected value of money you're spending given the number of stores-- but without telling you a specific number-- is whatever that number of stores turns out to be times the expected value of  $x$ . So this is a random variable that takes this as a numerical value whenever capital  $N$  happens to be equal to little  $n$ . This is a random variable, which by definition takes this numerical value whenever capital  $N$  is equal to little  $n$ .

So no matter what capital  $N$  happens to be what specific value, little  $n$ , it takes this is equal to that. Therefore the value of this random variable is going to be equal to that random variable. So as random variables, these two random variables are equal to each other.

And now we use the law of iterated expectations. The law of iterated expectations tells us that the overall expected value of  $y$  is the expected value of the conditional expectation. We have a formula for the conditional expectation. It's  $n$  times expected value of  $x$ .

Now the expected value of  $x$  is a number. Expected value of something random times a number is expected value of the random variable times the number itself. We can take a number outside the expectation.

So expected value of  $x$  gets pulled out. And that's the conclusion, that overall the expected amount of money you're going to spend is equal to how many stores you expect to visit on the average, and how much money you expect to spend on each one on the average. You might have guessed that this is the answer. If you expect to visit 10 stores, and you expect to spend \$100 on each store, then yes, you expect to spend \$1,000 today. You're not going to impress your Harvard friends if you tell them that story.

It's one of the cases where reasoning, on the average, does give you the plausible answer. But you will be able to impress your Harvard friends if you tell them that I can actually calculate the variance of how much I can spend. And we're going to work by applying this formula that we have, and the difficulty is basically sorting out all those terms here, and what they mean.

So let's start with this term. So the expected value of  $y$  given that you're visiting  $n$  stores is  $n$  times the expected value of  $x$ . That's what we did in the previous slide. So this thing is a random variable, it has a variance. What is the variance? Is the variance of  $n$  times the

expected value of  $x$ .

Remember expected value of  $x$  is a number. So we're dealing with the variance of  $n$  times a number. What happens when you multiply a random variable by a constant? The variance becomes the previous variance times the constant squared. So the variance of this is the variance of  $n$  times the square of that constant that we had here.

So this tells us the variance of the expected value of  $y$  given  $n$ . This is the part of the variability of how much money you're spending, which is attributed to the randomness, or the variability, in the number of stores that you are visiting. So the interpretation of the two terms is there's randomness in how much you're going to spend, and this is attributed to the randomness in the number of stores together with the randomness inside individual stores. Well, after I tell you how many stores you're visiting.

So now let's deal with this term-- the variance inside individual stores. Let's take it slow. If I tell you that you're visiting exactly little  $n$  stores, then  $y$  is how much money you spent in those little  $n$  stores. You're dealing with the sum of little  $n$  random variables. What is the variance of the sum of little  $n$  random variables? It's the sum of their variances.

So each store contributes a variance of  $x$ , and you're adding over little  $n$  stores. That's the variance of money spent if I tell you the number of stores. Now let's translate this into random variable notation. This is a random variable that takes this numerical value whenever capital  $N$  is equal to little  $n$ . This is a random variable that takes this numerical value whenever capital  $N$  is equal to little  $n$ . This is equal to that. Therefore, these two are always equal, no matter what happens to  $y$ .

So we have an equality here between random variables. Now we take expectations of both. Expected value of the variance is expected value of this. OK it may look confusing to think of the expected value of the variance here, but the variance of  $x$  is a number, not a random variable. You think of it as a constant. So its expected value of  $n$  times a constant gives us the expected value of  $n$  times the constant itself.

So now we got the second term as well, and now we put everything together, this plus that to get an expression for the overall variance of  $y$ . Which again, as I said before, the overall variability in  $y$  has to do with the variability of how much you spent inside the typical store. And the variability in the number of stores that you are visiting.

OK, so this is it for today. We'll change subjects quite radically from next time.