

ALBERT MEYER: In the last lecture, we spent time talking about the mean, or expectation, and its properties, most important one being linearity. But let's step back now and think about, what is it that the mean means? Why we care about it? We have this intuitive idea that if you do things long enough, if you keep experimenting with the same random variable collecting its values, its long run average will be about the same as its mean. Now, we're going to try to make that more precise.

So we're going to talk about the topic of deviation from the mean, or as I like to say, what does the mean really mean? Why do we care about it? Well, let's look at an example that's familiar to get a grip on the specific ideas that we're interested in. So suppose I toss a fair coin 101 times. Then, I know that the expected number, since all the values from zero through 101 are possible, and the middle value is the expectation, it's 50 and 1/2 heads. Now, I'm never going to get exactly 50 and 1/2 heads. The probability in 101 flips of getting 50 and 1/2 heads is zero because there's no way to flip 1/2 a head.

So you don't expect the expectation in that sense. No given measurement, no given experiment is going to yield the expectation. The expectation is this thing that we expect to come out on the average. Well, we can ask, what's the probability of getting as close as you could hope to get to the expectation? Namely, what's the probability of getting exactly 50 heads? And it's about 1/13. Or if you ask, what's the probability of getting either 50 or 51 heads, being within plus or minus one of the expectation? It's about 1/7.

OK, let's flip more coins and see what happens. This time I'm going to flip 1001 coins. And again, the expected number of heads is 500 and 1/2, which I'll never get exactly. The probability of getting exactly 500 heads is 1/39, and the probability of getting within one of the expectation, that is either 500 or 501 heads, is about 1/19.

Now, these numbers have gone down from the previous numbers. Remember, it was about 1/7 and we've gone down to 1/19. So it's actually we're less likely to be within a fixed distance, within one of the expectation when we flip more coins. So as the number of tosses grows, the number of heads gets less likely to be within any given fixed distance of the mean. But things get better when we start looking at percentages.

So what's the probability of being within 1% of the mean if I toss 1,001 coins? Well, 1% of

1,001 is about 10, so we're talking about 1% of the 1,001. And the probability of being within 10 of 500.5, that is to say the probability of being within 510 and 490, is about 0.49. It's almost 50-50, which is not really so bad. So we have a 50/50 chance of actually being within 1% of the expected number when I flip 1,001 coins.

So what we can start to say is that when we're trying to give the meaning to the mean, if I let μ be the standard abbreviation for expectation of R -- I'm doing that just so it'll fit on the slide nicely in formulas, so μ is the expectation of R -- the basic question we're asking is two basic questions. One is, what's the probability that the random variable is far from its mean, μ ? You could phrase that as, what's the probability that the distance from R to μ , the absolute value of R minus μ is greater than some amount, x . And the second question that we want to ask is, what's the average deviation? What's the expectation of the distance between R minus μ ? What's the expected value of r minus μ ?

Now, of course, we're trying to make sense of the meaning of the expectation, in terms of the expectation of the distance between R and this expectation. So there's a little bit of circularity there, but let's live with it and proceed.

Let's look at example to crystallize the ideas a little. Let's look at two dice with the same mean. The green die is going to be a standard fair die, in which each of the numbers one through six has an equal probability of showing up, and its expected value is exactly going to be the midpoint between one and six, or 3 and 1/2. Now, suppose I look at a loaded die, die two, which only throws a one or a six, but with equal probability. Then, its expectation is also 3 and 1/2, by the same reasoning.

So here are these two different die. One takes the values one through six equally likely, and the other takes only the two values one and six, but they have the same expectation. So how do I capture the difference? Well, if I look at the expected distance of the fair die to its mean, I claim it's one and a half. But the expected distance of the loaded die from its mean-- same mean remember, 3 and 1/2-- is actually 2 and 1/2. In fact, the second die is always exactly 2 and 1/2 from its expected value.

Let's look at the PDFs to get a grip on understanding what's going on. So here's the PDF for the fair die. Over one through six the probability is 1/6, so each of those green spikes, columns, is 1/6 high. And their total is the probability that the fair die takes one of those values one through six with equal likelihood.

Now, the expected value is exactly in the middle at 3 and 1/2. And the average distance of these points-- well, you can see that a third of the time, you're at distance 1/2, a third of the time, that is when you take the values 2 and 5, you are a distance exactly 1 and 1/2. And another third of the time, you're at distance 2 and 1/2 when you take one and six. And that averages out to the middle value of 1 and 1/2.

So the expected deviation, the expected distance, of the fair die from its mean is 1 and 1/2. On the other hand, for the loaded die, as we said, it's always exactly 2 and 1/2 from its expected value, which means its expected value is also 2 and 1/2. So we can start to see the difference between these two distributions and these two kinds of die. Even though they have the same expectation, one of them is more likely and has a greater expected distance from its mean.

And the moral of this short piece is that the mean alone is not a good predictor of a random variable's behavior, as you might suppose. One parameter, one number is not going to capture the shape of a PDF, which gives you more full information about the distribution of values of a random variable. And we need some more information than just the expectation. An especially, valuable extra piece of information that's still well less than the overall shape of the PDF of the random variable, is knowing its probable deviation from its mean.