Transcript

Live Video – Examples (A and B) – Part 1

01. 00:00 / 00:07 - Suppose 8% of US adults 20 years or older have diabetes. We can say that by saying the
02. 00:07 / 00:14 - prevalence of diabetes is 8% in our population of all US adults. And we wanted just think
03. 00:14 / 00:19 - about selecting some number I've left it open-ended right now. Select some number of people at
04. 00:19 / 00:23 - random, so I might say what if I'd like to select a hundred people at random. Well, then
05. 00:23 / 00:29 - how many would you expect to have diabetes if I selected a hundred? Eight. Right? Clearly
06. 00:29 / 00:33 - 8% of a hundred, very easy. We're going to learn a formula for that but you already know
07. 00:33 / 00:37 - how to do it without a formula. So remember that. What is the probability that none of
08. 00:37 / 00:41 - them will have diabetes? Or what's the probability that at least one of them will? And then we
09. 00:41 / 00:46 - might ask later what's the distribution of all possible things that could happen if I
10. 00:46 / 00:51 - were to do this over and over again. So if I sample 100 people today I might get exactly
11. 00:51 / 00:56 - eight, but I sample 100 people tomorrow I may get nine. I might even get four, sometimes
12. 00:56 / 01:01 - I might get none, and I may be lucky sample where nobody's diabetic or I can get it's
13. 01:01 / 01:06 - pretty rare I would imagine that I get all 100. It is possible, but very rare. We can
14. 01:06 / 01:10 - ask where is that distribution centered. Well I bet it is centered at about eight, because
15. 01:10 / 01:15 - that's what we would expect. And well, it's very likely for us to get seven or six or
16. 01:15 / 01:21 - four, nine, unlikely for us to get zero, extremely unlikely for us to get a hundred. We can even
17. 01:21 / 01:25 - already visualize some of what we think that picture would look like. Just from logic but
18. 01:25 / 01:29 - we'll be able to lay it down mathematically. Where is it centered, how spread out is it,
19. 01:29 / 01:33 - how much variability is there. We are asking the same kinds of questions that we asked
20. 01:33 / 01:38 - about variables when we start doing things that are a little bit more complicated. But
21. 01:38 / 01:43 - really if you think about it, X is just a variable. It's just a more complicated one
22. 01:43 / 01:49 - than what we've been thinking about. Age is a variable, but so is the number of diabetics
23. 01:49 / 01:54 - in a sample of size 100 if I do that repeatedly. It's a, quantity that varies. We'll have a
24. 01:54 / 01:58 - little bit more official definition later. But those are some other probabilities that
25. 01:58 / 02:05 - we can ask in this scenario that involve some aspect of probability. And then, here are
26. 02:05 / 02:09 - my comments about that the population is very large, we're looking at everybody in the whole
27. 02:09 / 02:15 - United States 20 or older. What that means if I pick one person at random, then I come
28. 02:15 / 02:20 - back and pick another person at random. They are completely disconnected, right they are,
29. 02:20 / 02:24 - what we will call independent observations knowing what happened to the first person
30. 02:24 / 02:29 - doesn't change the probability of what's going to happen on for the second person and so
31. 02:29 / 02:34 - on. We can make that assumption if they're making a really good random sample then what
32. 02:34 / 02:39 - that means is that these two observations are independent. As long as subjects are chosen in a
33. 02:39 / 02:45 - short period of time. Why does that matter? Well the prevalence of diabetes might be changing.
34. 02:45 / 02:49 - If I wait ten years will I might not even have the same probability anymore. But if
35. 02:49 / 02:54 - I'm just taking a sample today or this week or this month is pretty fixed, it's pretty
36. 02:54 / 02:59 - constant. If I sample people in a pretty short period it seems likely that the prevalence
37. 02:59 / 03:04 - will be the same for each selection. So the probability that one person has diabetes will
38. 03:04 / 03:10 - be eight percent every time I reach in and grab a person. We will learn we would calculate
39. 03:10 / 03:15 - probabilities here using the multiplication rule for Independents events, and later we'll
40. 03:15 / 03:20 - look at the binomial distribution. So, let's look at a similar question but on a small
41. 03:20 / 03:25 - population and ask ourselves some of these same questions. Suppose in a group of 60 subjects
42. 03:25 / 03:32 - maybe we're in a school or someplace where I have this fixed group of people. And I know
43. 03:33 / 03:38 - that five of them are diabetic. Then, I can ask some questions; if we select one person
44. 03:38 / 03:42 - at random this one you can answer right now, what is the probability that the person is
45. 03:42 / 03:47 - diabetic? Five out of sixty. Right? There's five total people that are diabetic I pick
46. 03:47 / 03:52 - one. So it's five chances to pick one who is diabetic out of sixty people total which
47. 03:52 / 03:59 - I think is like 1/12 or 0.083. I wanted to have it about the same eight percent, in both
48. 04:00 / 04:04 - examples. So if we select two subjects that at random what's the probably that both are
49. 04:04 / 04:09 - diabetic or none are diabetic. We'll come back and answer that at the end. And this

50. 04:09 / 04:14 - is again a little bit later of a topic if we select two subjects at random, what is
51. 04:14 / 04:19 - the expected number with diabetes? We can't answer that is easy as we could for the previous
52. 04:19 / 04:26 - scenario. We said 8% of 100 8. Easy! This is not independent because when I pull one
53. 04:27 / 04:32 - person out everything changes. Right? Now there are only four diabetics left if the
54. 04:32 / 04:36 - first person was diabetic or there are still five diabetics left if the first person wasn't.
55. 04:36 / 04:41 - The scenario is changing and in either case there's only fifty nine people left once I
56. 04:41 / 04:46 - pull somebody out. The whole situation is changing every time I pull somebody out this
57. 04:46 / 04:52 - hat this expected number is a little bit more difficult for us to contemplate. Here's my
58. 04:52 / 04:57 - comments about this one since the population is very small the result of one person will
59. 04:57 / 05:01 - change the subsequent probability so this means that they are dependent We will learn
60. 05:01 / 05:06 - that these probabilities can be calculated using conditional probability, especially
61. 05:06 / 05:10 - in the pick two case we are already saw it's easy to do the pick one case. We don't need
62. 05:10 / 05:14 - anything difficult to do that and the general multiplication rule.