Transcript

Live Video – Independence Part 1

01. 00:00 / 00:06 - Independent events, we're going to define it logically here now and then we'll formalize
02. 00:06 / 00:12 - the definition a little bit later. Two events are independent if knowing one event occurs
03. 00:12 / 00:18 - does not change the probability of the other. So if I picked two people from all people
04. 00:18 / 00:24 - in the United States, if they're picked randomly, the first persons answer doesn't change the
05. 00:24 / 00:30 - probability of what the second person is going to say. That's what we mean by independence.
06. 00:30 / 00:37 - Be careful this is not the same thing is disjoint. So in your mind English might tell you independent
07. 00:37 / 00:43 - means separate and that's great for English but it's totally not what statistics' definition
08. 00:43 / 00:50 - of independence is. Disjoint is events that are separate. Things that cannot happen together,
09. 00:50 / 00:56 - things that do not overlap, so disjoint is our word for that. Independence is a definition
10. 00:56 / 01:03 - related to whether or not the probability changes, knowing more information. So independence
11. 01:04 / 01:11 - is a statement about the sameness or the equality of the probability of one event whether or
12. 01:11 / 01:18 - not the other the other occurs. A simple example might be, let's say my two events you are
13. 01:18 / 01:25 - male and that you are football fan. So what this says is once I know you're a male what
14. 01:25 / 01:30 - does that tell me about the probability that you're a football fan? It's probably higher
15. 01:30 / 01:35 - than if I knew you were female. Right? Doesn't mean that you can't be a female football fan.
16. 01:35 / 01:40 - Right? Just said that my guess is that if I look at those two groups the probability
17. 01:40 / 01:45 - of being a football fan is higher among males than females. And doesn't that sound like
18. 01:45 / 01:50 - the kind of research that you might want to do in the real world. Not necessary football
19. 01:50 / 01:55 - fans. But that in this group, the probability is higher that this will happen than in the
20. 01:55 / 02:01 - other group. That's the kind of stuff that we want to be able to do. And so in the statistics
21. 02:01 / 02:07 - world we often are testing to see if two events are independent or not. If they're not independent
22. 02:07 / 02:11 - then they are dependent which means the probabilities are different. And then we want to know,
23. 02:11 / 02:15 - how different are they? Is it a big deal that men respond this way in women respond this
24. 02:15 / 02:20 - way to whatever medication I might be giving them, or whatever intervention I might be
25. 02:20 / 02:26 - doing? So those are kinds of questions that come into play about independence in statistics.
26. 02:26 / 02:33 - We are going to use this multiplication rule as a test. So this multiplication rule for
27. 02:33 / 02:40 - independent events says that if A and B are independent, then the probability of A and
28. 02:40 / 02:45 - B is equal to the probability A times B. So I have a question mark over this because what
29. 02:45 / 02:51 - I'm going to do is to see if this is true or not for our data. If I plug numbers into
30. 02:51 / 02:55 - the left hand side and the right hand side, and they are equal. Then the two of the events
31. 02:55 / 03:00 - are independent. If they're not equal then the two events are dependent. So this is one
32. 03:00 / 03:07 - way to test for independence. So the probability of A was 0.425, the probability of B was 0.329,
33. 03:09 / 03:16 - and the probability of A and B was 0.247. So if I multiply the probability of A times the
34. 03:17 / 03:24 - probability of B I get 0.1398 and then I want to ask, is that close, similar to my by hand
35. 03:24 / 03:30 - calculation not at all. And we are really know that these are dependent, knowing I have
36. 03:30 / 03:36 - calcium crystals present changes the probability that I'm going to have a high calcium
37. 03:36 / 03:41 - concentration or vice versa. It's more likely that I have a high calcium concentration in the yes
38. 03:41 / 03:48 - group than the no group. In data we can expect that four events that really are independent, we
39. 03:48 / 03:52 - get something that's slightly off. So once we come to statistical testing will be saying
40. 03:52 / 03:59 - is that far enough off that I know that it's true in my population. In like theoretical
41. 03:59 / 04:03 - probability "land" like flipping coins and experiments you'll get exactly. If they're
42. 04:03 / 04:07 - independent it will be exactly the same on the left and the right. But in data it'll
43. 04:07 / 04:14 - be close. You know, you cannot expect data come out exactly. But anyway so it's clearly
44. 04:14 / 04:20 - not equal not even close to being equal. So the question is, in the sample, for sure our
45. 04:20 / 04:25 - events are dependent even if it was off a smidge (tiny amount) technically in our sample
46. 04:25 / 04:29 - the results are slightly dependent. The question that we then ask is does that translate to
47. 04:29 / 04:35 - the population does this show the population is similarly dependent. Is this far enough
48. 04:35 / 04:40 - away for me to say this is definitely happening in the real whole population? So that's what

49. 04:40 / 04:46 - we'll be doing later. Right now, if I ask you to test it to see if events are independent
50. 04:46 / 04:50 - this, is one way, is that you need these three probabilities. And you multiply these two
51. 04:50 / 04:57 - together and see if it's equal to the probability that A and B. So it's just a test for independence.