

module_stat4_lecture6

📅 Thu, 2/17 4:18PM ⌚ 13:07

SUMMARY KEYWORDS

standard error, square root, estimate, quantitative variable, standard deviation, categorical variable, sample, proportion, fraction, average, number, support, data, respondents, candidate, equal, variance, divided, policy, remember

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

Since the last clip, I showed you how to do an estimate, and the standard error for a quantitative variable. So this clip, I'm going to begin by showing you the same thing. But if instead of a quantitative variable, you had a categorical variable. It's a categorical variable is one where it takes values one and zero. It's not numeric, it takes values in categories. So for example, do you support a particular candidate or not? Do you support a particular policy or not? Do you smoke or not? Are you single or not? Right? So those are all examples of categorical variables. And one may be interested in knowing what fraction of the population smoke, what fraction of the population support Candidate A, or policy P and things like that. So for a categorical variable, we're interested in estimating the fraction of the population with x equal to one. Okay, so the variable can either take value zero or one, we're interested in what fraction of the population has x equal to one. So the way to estimate this, is just look at the sample, right? And look at what fraction of the sample has that x equal to one. So that's my estimate of the proportion of the population, a fraction of the population who have x equal to one. So this is pretty simple. what else what else we're interested in is the standard error of the estimate as we had for a quantitative variable. And we use the same formula, it's s over square root of n . The only change now is that the standard deviation, S is the square root of P times one minus p . So it's this proportion that you estimated, so that p times one minus p , that's the variance. And the square root of that is the standard deviation. And if you plug that in here, you get the standard error of the estimate. So why this change? Right? So here? See, we don't have values like in the, in the quantitative variable, we have values like x_1, x_2, x_3 , so on, right? Like, think back of the example of sleepovers, it was A 10, A things like that, right? Here, we have either x equal to one, or x equal to zero. So it's a two possibility variable. So what we are in is the world of the Bernoulli random variable. Or in this particular case, because there are n people in the sample, this is like the binomial distribution. So if you think back to your probability, module three, we talked there about common discrete distributions, one of them was the binomial. Right? And, and this variance, it comes exactly from from there. So you may, if you weren't one, too, you may want to go back and check that out. So let me show you how to do it with a sample. An example. Right? So suppose you're interested what fraction support a particular policy? And let's say you collect a sample of size 10. And you asked this 10 people, do you support policy a or not? And you record the responses? These are people who

say, Yes, we support the people who say no, and so. So what you're interested in is estimating what fraction support the policy a right, so the firstly, the easiest one is the is this proportion to count the number of Y's so 1, 2, 3, 4, right, so 4 out of 10. So the estimate of P of this fraction in the population who support policy, that's going to be point. So straightforward. Now we want to calculate the standard error. So remember, the standard error is the square root of P times one minus p divided by square root of n, here n is 10. So the bottom is square root of 10. So this will be square root of term and on the top A P, we estimated it to be point four. So this will be point four times one minus b, that's point six. So that's square root of 24.24 divided by square root 10. And if you use a calculator, this becomes 0.15. So in this case, the standard error of the estimate is point 155. So this is how you would do example, sorry, this should have been categorical. Now, let me show you how you can do this also in Excel. Because you know, for a small data set, this is easy to do. But for large data sets, if you have 10, numbers, sorry, 100 numbers or so then this becomes a bit tedious. Okay, here, I have a Excel sheet with some data. So the first one is the sleep study again, right. And now let's say we have 12 respondents. So this is the respondent number one. And he or she says that the sleep eight hours, Respondent number two that aid number three, that's 10. And so So firstly, if we want to get the estimate of how many hours an average teenager sleeps, it's just going to be the average of these. So that's easy to calculate. And we have used this function before also in Excel. So it's just the average function. And from where this is from there, and we up to where it's up to you close the bracket. So the average, so that's my estimate. So the estimated number of hours and average teenager sleeps is 7.375 hours. And this is actually close to some studies, which have found that teenagers on average, the sleep between seven and seven hours in seven hour, 15 minutes. Now we want to next calculate the standard error. So firstly, we need to figure out the standard deviation. Remember, the standard error is the standard deviation divided by the square root of the number of observations, right. So to get the standard deviation, we use SD dV. Right. And again, it's going to be from here, up to here. So that's my standard deviation 1.53. And then to get the standard error, what I'm going to do is I'm going to take this, right, and divided by the square root of the number of observations. So here, the sample size is 12. So I'm going to divide it by the square root of 12. So my standard error is point 444. Here is another data. So this is the amount of student loans that his average student trying to find the amount of student loan that an average student in the US has upon graduation, right? So we collect a sample of 15 students, ask them about the debt levels at graduation, and these are the numbers we get. So now if I want to figure out what's the average debt level of an average student, right, so I'm going to do the average of these numbers. Right? So it'll be average from here, up to here. So that gives me it's about \$23,147. And again, this is close to what people have found researchers have found in the US the average student debt at graduation for the university students is about 25. Now again, to find the standard error, first, let's find out the standard deviation. So again, from here to here, closed bracket, so that's the standard deviation. And again, to find the standard error, I'm going to do standard deviation divided by the square root of n. Remember, they're 15 as the sample size here. So that's 15. So that's the standard error of test. So this is how you would do whenever you have quantitative variables like these are these. On the other side, if you're looking at categorical variables, that what fraction support a candidate A or Candidate B. So here you have a sample of 15. And what you have recorded is whether each person supports Candidate B, or candidate. So first thing for your estimate, remember, you need to figure out what fraction of these respondents or what proportion of this respondents support candidate. So the way to do that, you can even eyeball it, but what Excel allows you to use discount if, right, which gives you the count how many people support a among this 50. So firstly, you put in the range. And then you put in the criteria, which is within inverted commas. Sorry, within the inverted commas, your criteria is those who have said a. So if I hit this that says nine. So my proportion will be equal to this divided by the number of people in my sample, which is 15.

So that's point six. Now remember, for category for this type of data, the standard deviation is given by the square root of P times one minus p . So this case, it's point six times one minus P . That's point for this. And then to get the standard error, you just take the standard deviation, right, and divided by the square root of n . So here n is 15. So that's the standard error of interest.